



## Genome-scale comparison of expanded gene families in *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* with *Plasmodium malariae* and with other *Plasmodium* species <sup>☆</sup>



Hifzur Rahman Ansari <sup>a,1</sup>, Thomas J. Templeton <sup>b,c,1</sup>, Amit Kumar Subudhi <sup>a</sup>, Abhinay Ramaprasad <sup>a</sup>, Jianxia Tang <sup>d</sup>, Feng Lu <sup>d</sup>, Raece Naeem <sup>a</sup>, Yasmeen Hashish <sup>a</sup>, Mary C. Oguike <sup>e</sup>, Ernest Diez Benavente <sup>f</sup>, Taane G. Clark <sup>f,g</sup>, Colin J. Sutherland <sup>e,f,h</sup>, John W. Barnwell <sup>i</sup>, Richard Culleton <sup>j,\*</sup>, Jun Cao <sup>d,\*</sup>, Arnab Pain <sup>a,k,\*</sup>

<sup>a</sup> Pathogen Genomics Laboratory, Biological and Environmental Sciences and Engineering (BESE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Jeddah 23955-6900, Saudi Arabia

<sup>b</sup> Department of Protozoology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

<sup>c</sup> Department of Microbiology and Immunology, Weill Cornell Medical College, New York 10021, USA

<sup>d</sup> Key Laboratory of National Health and Family Planning Commission on Parasitic Disease Control and Prevention, Jiangsu Provincial Key Laboratory on Parasite and Vector Control Technology, Jiangsu Institute of Parasitic Diseases, Wuxi, Jiangsu Province, People's Republic of China

<sup>e</sup> Department of Immunology & Infection, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom

<sup>f</sup> Department of Pathogen Molecular Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom

<sup>g</sup> Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

<sup>h</sup> Public Health England Malaria Reference Laboratory, London School of Hygiene & Tropical Medicine, London, United Kingdom

<sup>i</sup> Centers for Disease Control and Prevention, Atlanta, GA 30329-4027, USA

<sup>j</sup> Malaria Unit, Department of Pathology, Institute of Tropical Medicine (NEKKEN), Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

<sup>k</sup> Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, N20 W10 Kita-ku, Sapporo 001-0020, Japan

### ARTICLE INFO

#### Article history:

Received 19 February 2016

Received in revised form 26 May 2016

Accepted 28 May 2016

Available online 5 July 2016

#### Keywords:

*Plasmodium ovale* spp.

*Plasmodium ovale wallikeri*

*Plasmodium ovale curtisi*

*Plasmodium malariae*

PIR

SURFIN

RBP-2

P25/27

### ABSTRACT

Malaria in humans is caused by six species of *Plasmodium* parasites, of which the nuclear genome sequences for the two *Plasmodium ovale* spp., *P. ovale curtisi* and *P. ovale wallikeri*, and *Plasmodium malariae* have not yet been analyzed. Here we present an analysis of the nuclear genome sequences of these three parasites, and describe gene family expansions therein. *Plasmodium ovale curtisi* and *P. ovale wallikeri* are genetically distinct but morphologically indistinguishable and have sympatric ranges through the tropics of Africa, Asia and Oceania. Both *P. ovale* spp. show expansion of the *surfin* variant gene family, and an amplification of the *Plasmodium* interspersed repeat (*pir*) superfamily which results in an approximately 30% increase in genome size. For comparison, we have also analyzed the draft nuclear genome of *P. malariae*, a malaria parasite causing mild malaria symptoms with a quartan life cycle, long-term chronic infections, and wide geographic distribution. *Plasmodium malariae* shows only a moderate level of expansion of *pir* genes, and unique expansions of a highly diverged transmembrane protein family with over 550 members and the gamete P25/27 gene family. The observed diversity in the *P. ovale wallikeri* and *P. ovale curtisi* surface antigens, combined with their phylogenetic separation, supports consideration that the two parasites be given species status.

© 2016 The Authors. Published by Elsevier Ltd on behalf of Australian Society for Parasitology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

<sup>☆</sup> Note: Sequence datasets reads were submitted to European Nucleotide Archive (ENA) under study accession numbers [PRJEB12678](https://www.ebi.ac.uk/ena/submit/study/PRJEB12678), [PRJEB12679](https://www.ebi.ac.uk/ena/submit/study/PRJEB12679) and [PRJEB12680](https://www.ebi.ac.uk/ena/submit/study/PRJEB12680) for *Plasmodium ovale curtisi*, *Plasmodium ovale wallikeri* and *Plasmodium malariae* isolates, respectively. Their assembly accession numbers are *P. ovale curtisi* ([GCA\\_900088555.1](https://www.ebi.ac.uk/ena/submit/assembly/GCA_900088555.1)), *P. ovale wallikeri* ([GCA\\_900088485.1](https://www.ebi.ac.uk/ena/submit/assembly/GCA_900088485.1), [GCA\\_900088545.1](https://www.ebi.ac.uk/ena/submit/assembly/GCA_900088545.1)) and *Plasmodium malariae* ([GCA\\_900088575.1](https://www.ebi.ac.uk/ena/submit/assembly/GCA_900088575.1)).

\* Corresponding authors at: Pathogen Genomics Laboratory, Biological and Environmental Sciences and Engineering (BESE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Jeddah 23955-6900, Saudi Arabia (A. Pain).

E-mail addresses: [richard@nagasaki-u.ac.jp](mailto:richard@nagasaki-u.ac.jp) (R. Culleton), [jipdjoe@163.com](mailto:jipdjoe@163.com) (J. Cao), [arnab.pain@kaust.edu.sa](mailto:arnab.pain@kaust.edu.sa) (A. Pain).

<sup>1</sup> Both authors contributed equally to this study.

## 1. Introduction

*Plasmodium ovale*, a causative agent of benign relapsing tertian malaria, was the last of the malaria parasite species infective to humans to be described (Stephens, 1922). Morphologically and phenotypically similar to the more common *Plasmodium vivax*, *P. ovale* is named for the typical oval appearance of infected erythrocytes. Its erythrocytic forms are, similar to those of *P. vivax*, restricted to reticulocytes, and infections typically result in low parasitaemias, cause mild clinical symptoms and are relatively short in duration (Faye et al., 1998). There are, however, occasional reports of more severe disease (Rojo-Marcos et al., 2008; Lau et al., 2013; Strydom et al., 2014; Tomar et al., 2015). *Plasmodium ovale* shares with *P. vivax*, amongst the malaria parasites infective to humans, the ability to produce hypnozoites, dormant liver stages that can cause relapse months or even years following initial exposure to sporozoites. These forms constitute a barrier to the control of *ovale* and *vivax* malaria, as they are treatable by only a single US Food and Drug Administration (FDA) approved drug, primaquine.

Two morphologically identical but genetically distinct forms of *P. ovale* have been identified through the typing of genetic markers (Tachibana et al., 2002; Winn et al., 2004). These have been argued to constitute separate species, as the absence of hybrid forms suggests they are reproductively isolated, although their ranges are sympatric and occasionally infect the same host (Sutherland et al., 2010; Fuehrer et al., 2012). Since the two *P. ovale* forms were first described and assigned the sub-species trinomials *P. ovale curtisi* and *P. ovale wallikeri* (Sutherland et al., 2010), compelling evidence has accrued that the two variants are, indeed, separate species (Fuehrer et al., 2012; Fañcony et al., 2012; Putaporntip et al., 2013; Oguike and Sutherland, 2015).

*Plasmodium ovale* spp. are found predominantly in Sub-Saharan Africa, but are also present on the islands of the Western Pacific and, sporadically, in southeastern Asia and India (Lysenko and Beljaev, 1969; Oguike et al., 2011; Bauffe et al., 2012; Fuehrer et al., 2012; Fañcony et al., 2012; Putaporntip et al., 2013; Chaturvedi et al., 2015). Frequently occurring in mixed infections with other malaria parasite species, most notably *Plasmodium falciparum* and *Plasmodium malariae*, the prevalence of *P. ovale* spp. is most certainly underestimated, as its typical low parasitemia and predilection for co-infection render it difficult to diagnose by light microscopy (Fañcony et al., 2012; Dinko et al., 2013). This is likely to be especially true in high malaria transmission regions such as West Africa, where other malaria parasite species are common and immunity against malaria is high in local populations. Sensitive PCR detection methods indicate a prevalence of *P. ovale* spp. of 14% and above in parts of Africa (May et al., 1999; Tobian et al., 2000), versus estimates based on microscopy which are typically below 5% (Mueller et al., 2007). Thus the public health impact of *ovale* malaria is probably underestimated.

With a range that closely mirrors that of *P. falciparum*, the quartan malaria parasite *P. malariae* (Grassi and Feletti, 1892) has a widespread, although patchy distribution (Garnham, 1966). *Plasmodium malariae* causes a relatively mild and chronic disease, characterised by a fever periodicity of 72 h, in contrast to the 48 h cycles of all other human malaria parasites. During its erythrocytic cycle it shows a strong invasion preference for mature red blood cells. As with the *P. ovale* spp., its prevalence is difficult to gauge accurately, as its presence is often masked due to co-infection with *P. falciparum*.

Comparative genomics allows the investigation of the genetic determinants of phylogenetic differences between species. For malaria parasites, comparative studies can help to understand the diversity of *Plasmodium* adaptations to infection of host tissues, the resulting spectra of disease and the evolutionary processes that

have led to speciation. Comparative genomic studies of *P. ovale curtisi* and *P. ovale wallikeri* with other malaria parasite species are of interest due to their characteristically mild infections and ability to cause relapse through the activation of dormant hypnozoite stages in the liver. Similarly, the genome of *P. malariae* may hold clues to understanding the regulation of periodicity. To date, the genome sequencing of *P. ovale* has been hindered by the lack of adequate concentrations of parasite DNA, a consequence of the low parasite densities characteristic of this species, and prevalence of mixed infections with other *Plasmodium* spp. This difficulty has diminished with the advent of whole genome sequencing technologies that require smaller starting concentrations of genomic DNA and high sequencing coverage overcomes problems with high levels of contaminating host DNA.

Here, we report the genome sequence and analysis of two isolates of *P. ovale wallikeri*, both obtained from infected migrant workers returning to China from West Africa; and two isolates of *P. ovale curtisi*, one from a Chinese worker returning from West Africa and the chimpanzee-propagated US Centers for Disease Control and Prevention (CDC) Nigeria I strain (Collins et al., 1987). For detailed comparison, we have derived a partial genome of *P. malariae* from the CDC Uganda I strain (Collins et al., 1984, 1990). Phylogenetic analyses support a close relationship but separation of *P. ovale wallikeri* and *P. ovale curtisi*, and their grouping with *P. malariae*, *P. vivax* and related non-human primate malaria parasites. In the context of observations from the wider *Plasmodium* genus we present an analysis of variant antigen gene families encoding invasion proteins, and parasite proteins predicted to be exported to the surface membranes of the parasite or infected erythrocyte. The implications for the species status of these two parasites are considered.

## 2. Materials and methods

### 2.1. Parasite materials

Blood samples of two isolates of *P. ovale wallikeri* (designated as *P. ovale wallikeri* 1, Pow1; and *P. ovale wallikeri* 2, Pow2) and one isolate of *P. ovale curtisi* (designated as *P. ovale curtisi* 2, Poc2) were obtained from febrile patients, Chinese males with ages between 23 and 45, presenting at local hospitals in Jiangsu Province, China, after returning from labouring work in African countries. Diagnosis of *P. ovale* was determined by microscopy and confirmed by PCR, typing to *P. ovale wallikeri* (Gabon, two samples) and one isolate of *P. ovale curtisi* (Nigeria). The real-time TaqMan PCR (TqPCR) assay that targets the *ssrRNA* gene was carried out to detect *P. ovale* sub-species according to previous publication (Bauffe et al., 2012). Briefly, amplification was performed using the following sets of primers: POF: 5'-ATAAACTATGCCGACTAGGTT-3' and POR 5'-ACTTTGATTCTCATAAGGTA-3', for the detection of *P. ovale wallikeri* probe pPOW HEX-AATTCCTTTGGAAATTTCTAGATTG-B HQ1 was used and probe pPOC FAM-TTCCTTCGGGGAA-ATTTCTTAGA-BHQ1 was used for *P. ovale curtisi* detection. The real-time PCRs were carried out using Light Cycler TaqMan master (Roche, Germany) on Roche LightCycler® 480 (Roche). Following informed written consent (approved by the Institutional Review Board of Jiangsu Institute of Parasitic Diseases, China, IRB00004221) five ml of whole blood were collected prior to treatment for malaria. White blood cells (WBCs) were removed by passing through a non-woven fabric (NWF) filter (Tao et al., 2011) and genomic DNA was extracted using a QIAamp DNA blood mini kit according to the manufacturer's instructions (Qiagen, Germany). Samples were stored in a DNAstable tube (Biomatrica, USA) until use. The *P. malariae* Uganda I CDC strain was isolated in 1982 from

**Table 1**  
Genome assembly and completeness statistics of five *Plasmodium* malarial species.

	<i>P. ovale wallikeri</i> _1 (Pow1)	<i>P. ovale wallikeri</i> _2 (Pow2)	<i>P. ovale curtisi</i> _1 (Poc1, Nigeria I)	<i>P. ovale curtisi</i> _2 (Poc2)	<i>P. malariae</i> (Pmal, Uganda I)
Number of scaffolds ( $\geq 1000$ bp)	1,378	1,593	4,025	2,227	7,270
Total length (bp)	35,260,179	35,103,764	34,518,662	38,010,039	31,925,186
N50 <sup>a</sup>	174,450	136,984	17,984	46,061	6398
GC (%)	28.91	29.12	28.46	27.76	24.74
Number of protein coding genes	8582	8813	7950	8825	6462
Number of complete BUSCOs <sup>b</sup> (429)	344	342	344	369	329
Number of Fragmented BUSCOs	29	32	27	24	37

<sup>a</sup> N50 represents a number where 50% of the entire assembly was contained in contigs or scaffolds equal to or larger than this value.

<sup>b</sup> The completeness of assemblies was determined by the BUSCO tool (Simão et al., 2015).

an infant given a blood transfusion from an individual who had immigrated to the United States from Uganda in 1974. The Nigeria I CDC strain of *P. ovale* was isolated in 1977 from a member of the Peace Corps stationed in Nigeria and was genotyped as *P. ovale curtisi* (herein referred to as *P. ovale curtisi* 1, Poc1). Both the Uganda I strain of *P. malariae* and the Nigeria I strain of *P. ovale curtisi* were maintained in chimpanzees under protocols established in accordance with the rules and regulations of the U. S. Animal Welfare Act. WBCs and platelets were removed from infected chimpanzee blood collected in heparin by passage of adenosine diphosphate (ADP) activated blood sequentially through 0.1 mm acid washed glass beads, Plasmodipur filters, and a column of Whatman CF11 cellulose fibers. Erythrocytes infected with late blood stage parasites were concentrated and collected from Percoll density cushions and extracted using Qiagen blood kits.

## 2.2. Genome sequencing and assembly

Sequencing libraries were prepared from the DNA samples from *P. ovale curtisi* (Poc1) and *P. malariae* Uganda I/CDC using the Diagenode MicroPLEX V2 protocol ([www.diagenode.com](http://www.diagenode.com)), following the manufacturer's instructions. For the *P. ovale wallikeri* (Pow1 and Pow2) and *P. ovale curtisi* (Poc2) samples, genomic DNA in 50  $\mu$ l of elution buffer was sheared using Covaris E220 using the following parameters: 175 W peak incident power, 5% duty factor, 200 cycles per burst and 65 s treatment time to obtain 350 bp target fragment size, and 25 s treatment time to obtain 550 bp target fragment size, according to manufacturer's protocol (<http://covarisinc.com/products/afa-ultrasonication/e-series/>). DNA library was prepared using a TruSeq DNA PCR-Free Library Preparation Kit (Illumina, USA) followed by 10 PCR amplification cycles using a KAPA HiFi Hotstart PCR kit (Kapabiosystems, South Africa). The prepared libraries were subjected to routine quality control checks and quantification before several rounds of Illumina MiSeq (600 cycles kit) and HiSeq2000 (101 bp paired-end) runs to yield high sequence coverage with read lengths of 300–600 bp and 97% of the reads having quality scores greater than or equal to Q20. Sequencing adaptors were removed by the Trimmomatic tool (Bolger et al., 2014). Trimmed reads were bioinformatically screened for host contamination, based on the respective vertebrate hosts. The filtered reads were assembled using the A5-miseq pipeline (Coil et al., 2015). Assemblies were scaffolded (SSPACE), gap closed (IMAGE) and error corrected with ICORN (Swain et al., 2012) and Pilon (Walker et al., 2014). Scaffolds were ordered by ABACAS (Assefa et al., 2009) and Artemis (Rutherford et al., 2000), using published *P. vivax* Salvador I as a reference strain. Since infections with mixed parasite genotypes are common in *P. ovale* and *P. malariae* natural infections, we verified the number of predominant genotypes in each isolate by manually checking (data not shown) at least eight polymorphic genes (MSP1,

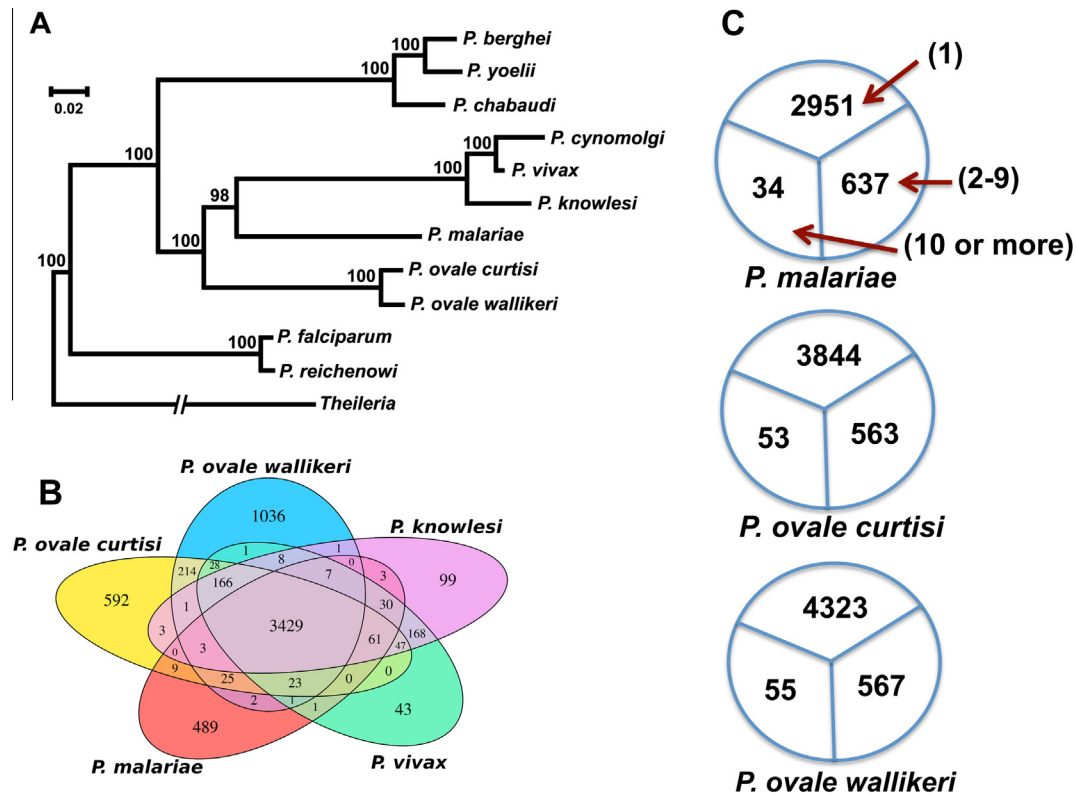
AMA1, CSP, Pfs230, p12, RAD51, g377 and CCp1) with aligned Illumina reads and looking for single nucleotide polymorphisms (SNPs) using the BamView option in Artemis (Carver et al., 2013). These genes were selected to confirm clonality and identity of the samples. In addition a whole genome scan for multiplicity of infection was performed using the estMOI tool (Assefa et al., 2014) and all of the genes tested also were confirmed as a single genotype (Supplementary Table S1). The basic assembly and annotation statistics are provided in Table 1.

## 2.3. Phylogenetic analyses

A phylogenetic tree for *Plasmodium* was generated by first selecting a set of 460 1:1 ortholog clusters, derived from OrthoMCL output (Li et al., 2003) (amino acid length greater than or equal to 200) from all sequenced *Plasmodium* spp. that are shown in Fig. 1A. Sequences were concatenated after multiple sequence alignment and trimming by MUSCLE (Edgar, 2004) and Trimal (Capella-Gutiérrez et al., 2009), respectively, to give a concatamer of 127,351 amino acid length. MrBayes (Ronquist et al., 2012) was used for phylogenetic tree construction with a fixed-rate Jones amino acid model and 40,000 generations. In a parallel approach, we also prepared a manually curated set of amino acid concatamers comprised of aligned amino acid sequences from selected ribosomal proteins, surface antigens, glycolytic enzymes, elongation factors and proteins encoded on the mitochondrial and apicoplast genomes (Supplementary Table S2). This concatamer was 7,532 amino acids long. Orthology of the included proteins was manually assessed for conserved synteny, and supported by best-hit sequence similarity in BLAST screens. We used orthologous sequences from *Theileria annulata* as an out-group and phylogenetic relationships of 10 *Plasmodium* spp. including both *P. ovale* spp. and *P. malariae* were inferred using the TreeGraph2 tool (Stöver and Müller, 2010).

## 2.4. Gene prediction and genome annotation

Genes were identified by Augustus v3 (Stanke et al., 2006), which was trained on a combined set of 374 manually annotated *Plasmodium knowlesi* and *P. vivax* genes from GeneDB (Logan-Klumpler et al., 2012) and a small subset of *P. ovale* genes obtained by manual curation with the help of BLAST similarity between predicted *P. knowlesi* strain H and *P. vivax* Sal1 genes. The same set of genes was also used as a training set for annotation of *P. malariae*. Ortholog information amongst other *Plasmodium* spp. was obtained by OrthoMCL (Li et al., 2003). Annotation of the *P. ovale wallikeri* and *P. ovale curtisi* genomes was performed using local BLASTP screens of their predicted proteomes using select queries from other *Plasmodium* spp. with manually curated annotation available in Gene DB (Logan-Klumpler et al., 2012). Such *P. ovale* spp. proteins were then used as BLAST queries either against the



**Fig. 1.** Comparison of *Plasmodium ovale wallikeri*, *Plasmodium ovale curtisi* and *Plasmodium malariae* with respect to other sequenced *Plasmodium* spp. using phylogeny and gene-orthology. (A) Phylogenetic tree showing the position of *P. ovale wallikeri*, *P. ovale curtisi* and *P. malariae* with respect to other sequenced *Plasmodium* spp. The Bayesian tree was generated by first selecting 460 1:1 ortholog clusters (amino acid length  $\geq 200$ ). Sequences were concatenated after multiple sequence alignment and trimmed by MUSCLE (Edgar, 2004) and Trimal (Capella-Gutiérrez et al., 2009), respectively, resulting in a concatenation of 127,351 amino acids. The MrBayes program was used for phylogenetic tree construction with an amino acid model and 40,000 generations. The numbers at the nodes indicate their probability and branch lengths are proportional according to the scale given. (B) Venn diagram of orthologous groups (as derived by Ortho-MCL (Li et al., 2003)) indicates the shared and unique gene clusters across five malaria species which infect humans. (C) Number of total orthologous groups with only one, two to nine, and more than 10 members in the respective groups, as derived by OrthoMCL for *P. ovale wallikeri*, *P. ovale curtisi* and *P. malariae*.

*P. ovale* spp. predicted proteomes, *Plasmodium* predicted proteomes at The National Center for Biotechnology Information (NCBI) GenBank using PSI-BLAST, or select *Plasmodium* genomes using the BLAST application at PlasmoDB ([www.Plasmodb.org](http://www.Plasmodb.org)). Additional annotation tools included signal peptide (SignalP 4.1; [www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/)), transmembrane prediction (TMHMM ver 2.0; [www.cbs.dtu.dk/services/TMHMM/](http://www.cbs.dtu.dk/services/TMHMM/)), NCBI conserved domain predictor ([www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi](http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi)) and InterPro scan ([www.ebi.ac.uk/interpro/](http://www.ebi.ac.uk/interpro/)). Hidden Markov Models (HMMs) that were used to identify individual members of all expanded gene families in both *P. ovale* and *P. malariae* are provided in Supplementary Raw Data S1. The presence of a canonical PEXEL motif in protein sequences was determined using ExportPred v2.0 with a cut-off value of 1.5 (Boddey et al., 2013) and the presence of a PEXEL-like motif was determined using the Sequence Manipulation Suite (Stothard, 2000).

#### 2.5. Identification of SNP and insertions/deletions (indels)

The raw Illumina reads of Poc1 and Poc2 were quality trimmed and mapped against Pow1 and Pow2 references using bwa (Li and Durbin, 2009). The resulting alignments were pre-processed to remove PCR duplicates using picard (<http://broadinstitute.github.io/picard/>). The alignments were re-aligned and base calibrated, variants were identified using the GATK HaplotypeCaller program (McKenna et al., 2010). High confidence SNPs and indels were obtained by applying a quality filter as recommended in the GATK

best practices document. The SNP annotation was added by using the SnpEff tool (Cingolani et al., 2012).

### 3. Results and discussion

#### 3.1. Genome assembly and construction of orthologous groups

The nuclear genomes of both *P. ovale wallikeri* isolates assembled at approximately 35 Mbp genome size, whereas there was a difference in the assembled genome sizes of the *P. ovale curtisi* isolates. The discrepancy in the draft genome sizes between *P. ovale curtisi* 1 and *P. ovale curtisi* 2 is primarily due to differences in the magnitude of the repertoire of the assembled *Plasmodium* interspersed repeat (*pir*) gene family. Taking into account these discrepancies, both *P. ovale* spp. genomes show a minimum of 30% increase in their nuclear genome sizes owing to expansion of variant gene families. Manual inspection of candidate polymorphic antigen genes (e.g. MSP1, AMA1, Pfs230 etc) and a genome-wide analysis by the estMOI tool (Assefa et al., 2014) verified the presence of a predominant single genotype in each of the sequenced isolates (Supplementary Table S1), thus greatly reducing the possibility that the increase in genome size is due to the sequencing of a multi-clone infection. The *P. malariae* nuclear genome does not show any evidence of high levels of *pir* gene amplification, beyond what has been observed in published *Plasmodium* genomes (see, for example, Tachibana et al., 2012). We assessed the completeness of the predicted genomes using the BUSCO tool (Simão et al., 2015) and concluded that both *P. ovale* spp. are more complete (80–86%

**Table 2**  
Expanded gene families of *Plasmodium ovale* and *Plasmodium malariae* and comparison with other *Plasmodium* parasites.

	Number of Genes									Comments
	<i>P. ovale</i>				other <i>Plasmodium</i> spp.					
	Pow1	Pow2	Poc1	Poc2	<i>P. mal</i>	<i>P. fal</i>	<i>P. viv</i>	<i>P. cyn</i>	<i>P. kno</i>	
<i>PIR</i> multigene family	>1800 <sup>a</sup>	>1800 <sup>a</sup>	>1900 <sup>a</sup>	>2100 <sup>a</sup>	>250 <sup>a</sup>	>200 <sup>b</sup>	~1200 <sup>b</sup>	265 <sup>c</sup>	71 <sup>b</sup>	<i>Plasmodium</i> interspersed repeat ( <i>pir</i> ); in <i>P. vivax</i> termed <i>vir</i> , in <i>P. ovale</i> termed <i>oir</i> , in <i>P. malariae</i> termed <i>mir</i>
<i>STP1/Surfin</i>	133 <sup>a</sup>	141 <sup>a</sup>	59 <sup>a</sup>	80 <sup>a</sup>	129 <sup>a</sup>	10 <sup>a, b</sup>	2 <sup>b</sup>	3 <sup>b</sup>	0 <sup>b</sup>	Termed <i>PvSTP1</i> (sub-telomeric protein 1) in <i>P. vivax</i> .
<i>tryp-rich antigen</i>	41 <sup>a</sup>	39 <sup>a</sup>	39 <sup>a</sup>	39 <sup>a</sup>	33 <sup>a</sup>	4 <sup>a, b</sup>	36 <sup>c</sup>	36 <sup>c</sup>	26 <sup>b</sup>	Termed <i>Pv-fam-a</i> in <i>P. vivax</i>
<i>phist</i>	50 <sup>a</sup>	50 <sup>a</sup>	55 <sup>a</sup>	59 <sup>a</sup>	31 <sup>a</sup>	74	n.d.	n.d.	n.d.	Termed <i>Pv-fam-e</i> and <i>rad</i> in <i>P. vivax</i>
<i>RBP (all)</i>	12 <sup>a</sup>	12 <sup>a</sup>	13 <sup>a</sup>	12 <sup>a</sup>	5 <sup>a</sup>	6 <sup>b</sup>	8 <sup>b</sup>	8 <sup>c</sup>	2	Reticulocyte binding proteins (RBPs). Numbers in bracket represent putative pseudo genes in each RBP sub-family due to presence of in-frame stop codons.
<i>RBP1</i>	1	1	2	2	2(1)	1	2	1	n.d.	
<i>RBP2</i>	9(5)	9(5)	9(2)	8(3)	2(1)	2	5	5	n.d.	
<i>RBP3</i>	2(1)	2(1)	2(1)	2(1)	1	1(1)	1(1)	1	n.d.	
<i>RBP4(rh4)</i>	0	0	0	0	0	1	0	0	n.d.	
<i>RBP5(rh5)</i>	0	0	0	0	0	1	0	0	n.d.	
Ring-exported protein (REX)	6	6	5	5	1	4 <sup>b</sup>	2 <sup>b</sup>	0 <sup>b</sup>	1 <sup>b</sup>	Maurer's cleft protein Hawthorne et al. (2004)
P28	4	4	3	3	1	1 <sup>b</sup>	1 <sup>b</sup>	0 <sup>b</sup>	1 <sup>b</sup>	Ookinete surface protein
6-cys family	15 <sup>a</sup>	15 <sup>a</sup>	13 <sup>a</sup>	14 <sup>a</sup>	12 <sup>a</sup>	13	13	n.d.	n.d.	Examples include P48/45 and P230 ookinete surface antigens, and list includes the 6-cys-like protein, SSP3
KELT	8	8	8	8	1	1	1	1	1	Telomeric gene in <i>P. falciparum</i> . Expanded in tandem array in <i>P. ovale</i> . Signal peptide in <i>P. ovale</i> versions.
Gamete antigen 27/25	0	0	0	0	16	1	0	0	1	RNA binding phosphoprotein. Abundant in early gametocyte stage.
Pm-fam-a gene family	17 <sup>a</sup>	13 <sup>a</sup>	12 <sup>a</sup>	12 <sup>a</sup>	>550 <sup>a</sup>	0	21	11	5	Novel gene family identified in <i>P. malariae</i> . >75% genes in this family in <i>P. malariae</i> code for predicted 2-transmembrane domains. >350 genes contain domain of unknown function DUF3671 (PF12420) of which >200 are <i>Pv-fam-b</i> like proteins.

Pow, *P. ovale wallikeri*; Poc, *P. ovale curtisi*; *P. mal*, *P. malariae*; *P. fal*, *Plasmodium falciparum*; *P. viv*, *Plasmodium vivax*; *P. cyn*, *Plasmodium cynomolgi*; *P. kno*, *Plasmodium knowlesi*., n.d., not determined.

<sup>a</sup> Pseudo genes, truncated genes and gene fragments are included.

<sup>b</sup> Data is from GeneDB.org and PlasmoDB -v26 ([www.plasmodb.org](http://www.plasmodb.org)).

<sup>c</sup> Data is from Tachibana et al. (2012).

BUSCO coverage) and hence more representative of their species than *P. malariae* (76%) in our datasets. We constructed a comparative orthology map of these three *Plasmodium* genomes together with the other three human malaria parasite species (*P. falciparum*, *P. vivax* and *P. knowlesi*) and identified the distribution of shared orthologs and gene families by the OrthoMCL tool. We identified a core group of 3,429 1:1 orthologs present across human malaria parasite species (Fig. 1B, Supplementary Table S3). The relevant genome assembly and gene prediction statistics are provided in Table 1. The observations of unusually large genome sizes, expansion of gene families and genomic variations were independently observed and verified in two isolates of natural infections for *P. ovale wallikeri*, and one natural and one laboratory strain for *P. ovale curtisi*. To the best of our knowledge, the assemblies reported here and the corresponding automated annotation data are trustworthy. However, it is important to note that all such automated prediction datasets need further experimental validation before error-free assemblies and precisely annotated gene boundaries may be confirmed for future follow-up experiments.

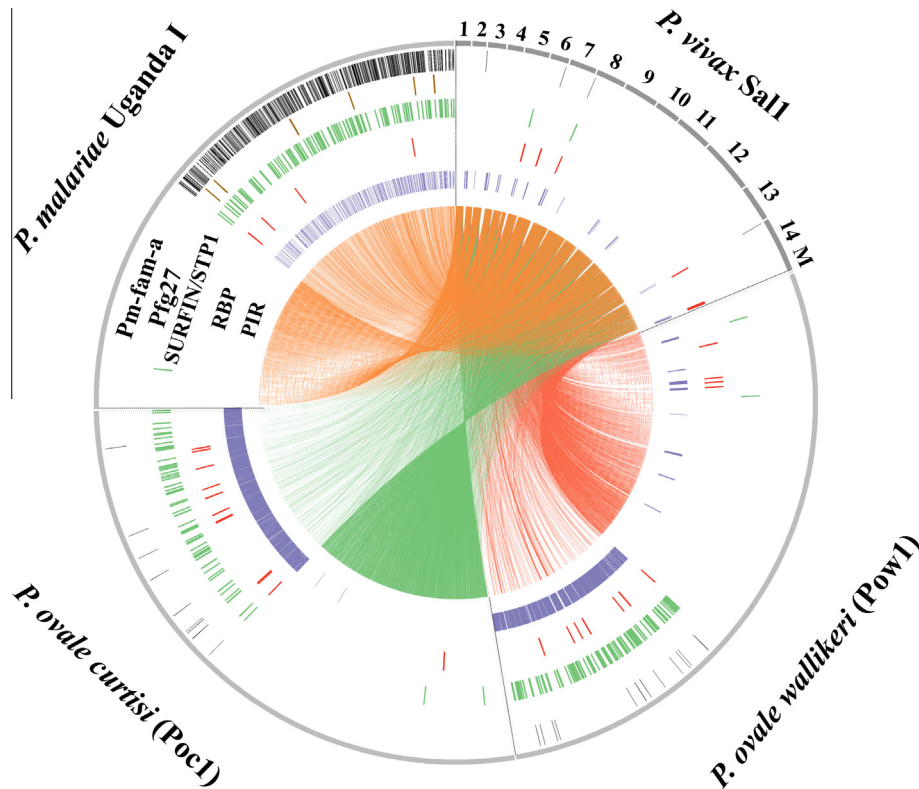
### 3.2. Phylogenetic placement of *P. ovale* spp. and *P. malariae* within *Plasmodium*

Phylogenetic trees for *Plasmodium* were derived using amino acid concatamers comprised of either OrthoMCL outputs or manually selected and curated proteins (see Section 2.3 for description

of the concatamers). The trees similarly showed that the *ovale* parasites group together, and more broadly in a clade with *P. malariae*, *P. vivax*, *P. knowlesi* and *Plasmodium cynomolgi*, to the exclusion of the clades encoding rodent malaria parasites, as well as the clade of *P. falciparum* plus *Plasmodium reichenowi* (Fig. 1A). The phylogeny confirms with strong probability and support the recently published trees inferred using mitochondrial, apicoplast and amino acid concatamer sequence information (Sutherland et al., 2010; Schaefer et al., 2013), and contrasts with a study describing a relationship between *P. ovale* spp. and rodent malaria parasites (Arisue et al., 2012). The following observations on the distribution of genes and their degree of amplification is considered in the context that *P. ovale* is related to *P. vivax* and non-human primate malaria parasites. Examples of discussed proteins are listed in Table 2, Supplementary Tables S4–S7.

### 3.3. Amplified protein families: SURFIN

Similar to other members of *Plasmodium*, the *P. ovale curtisi*, *P. ovale wallikeri* and *P. malariae* genomes harbour numerous multi-gene families that encode proteins secreted to the parasite surface membrane or exported to the infected erythrocyte cytoplasm and surface. Patterns of expansion of these families in *P. ovale* spp. were compared with those in the human malaria parasites *P. vivax* and *P. falciparum* (Table 2, Fig. 2). Distinctions in expansion include the *surfin* (termed *Pvstp1* in *P. vivax*) gene family which is found across



**Fig. 2.** Gene family expansion in *Plasmodium ovale wallikeri*, *Plasmodium ovale curtisi* and *Plasmodium malariae*. Circos plot (Krzywinski et al., 2009) showing shared regions (inner most colored ring) between *P. ovale wallikeri* (red), *P. ovale curtisi* (green) and *P. malariae* (brown) with *P. vivax*, derived by PROMER tool (Kurtz et al., 2004) mapping. Expanded gene family locations in our assembled contigs are marked as (from inner most to outside ring), Purple, PIR; Red, reticulocyte binding proteins (RBPs); Green, STP1/SURFIN; Brown, Pfg27; and Black, Pm-fam-a gene family.

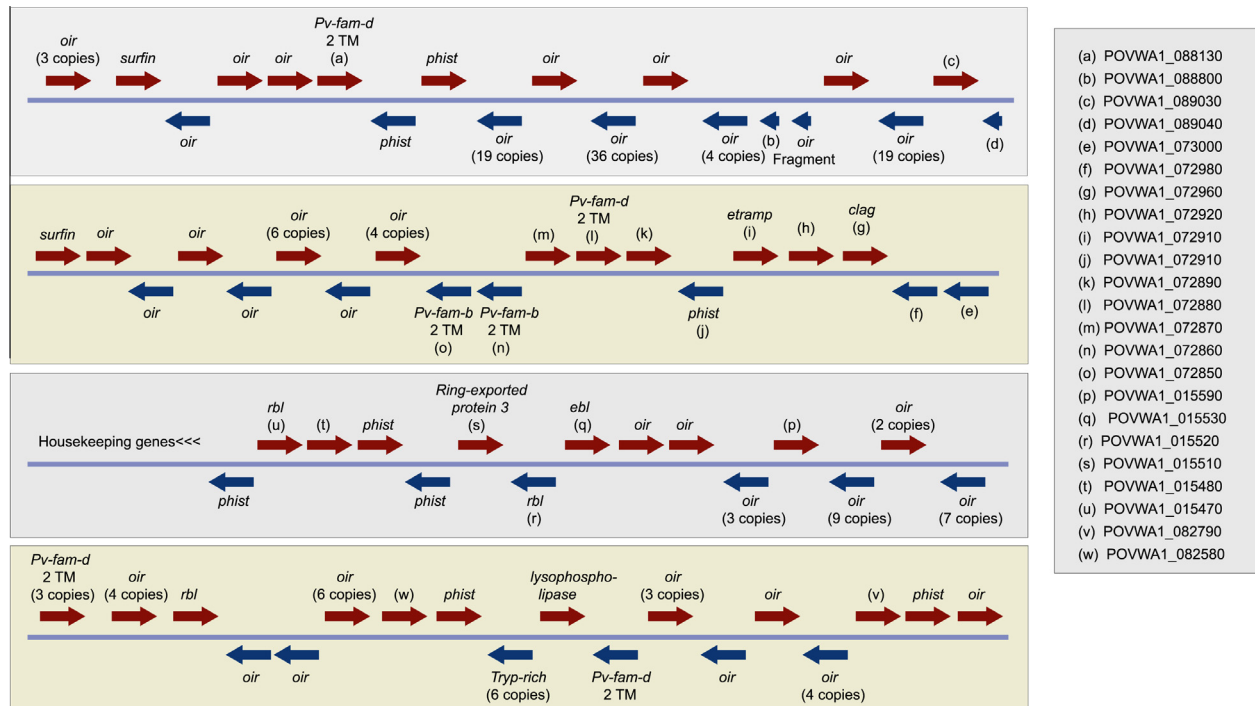
*Plasmodium*, with the exception of rodent malaria parasites, but is amplified to a much greater degree in *P. ovale* spp. Specifically, over 50 *surfin* genes are present in *P. ovale curtisi* while we observed more than 125 genes in *P. ovale wallikeri* and *P. malariae*, indicating a differential expansion (Table 2). In contrast, 10 *surfin* genes are present in *P. falciparum* and only two in *P. vivax*.

The *surfin* genes encode transmembrane (TM) proteins, which are exported to the host erythrocyte cytoplasm in a manner independent of either signal peptide or PEXEL/HT motif, and are thought to be exposed on the erythrocyte surface (Winter et al., 2005; Alexandre et al., 2011; Zhu et al., 2013). The cytoplasmic region is more highly conserved between paralogs and contains multiple copies of a tryptophan-rich domain whose function is not known and is also found in the SICAvir proteins of *P. knowlesi* (Winter et al., 2005). Two other *Plasmodium* exported products, the tryptophan-rich antigen and PHIST domain proteins, also have hallmarks of abundant, conserved tryptophan residues; although it is not known whether this indicates a shared structural or functional feature, such as stabilisation of globular domains or participation in hydrophobic interactions with other proteins or lipid membranes. Genes encoding PHIST domain proteins are amplified in all human malaria parasites, with over 50 genes in both *P. ovale curtisi* and *P. ovale wallikeri*. Similarly, the tryptophan-rich antigen genes (*tra*, described in Section 3.6) are also greatly expanded in *P. ovale curtisi*, *P. ovale wallikeri* and *P. malariae* as in other *Plasmodium* spp.

#### 3.4. Amplified gene families: PIR

*Plasmodium vivax*, primate and rodent malaria parasites differ from *P. falciparum* in the presence of a highly amplified gene fam-

ily, termed *vir* (*vivax interspersed repeat*, from *P. vivax*; or generally, *pir*, for *Plasmodium interspersed repeat*; del Portillo et al., 2001a,b; Cunningham et al., 2010; Lopez et al., 2013) with a nomenclature based upon the species name (eg. *kir* in *P. knowlesi* and *bir* in *Plasmodium berghei*). Similar to the SURFIN proteins, the *pir* gene products are predicted to be exported to the infected erythrocyte surface, and lack typical signal peptide sequences and PEXEL/HT erythrocyte trafficking motifs. However, approximately 160 VIR proteins contain PEXEL-like motifs for the purpose of protein export to the host cell surface (Lopez et al., 2013). The proteins possess TM domains and short cytoplasmic domains less than 60 amino acids in length. The extracellular domain shows homology to the SURFIN extracellular domain, and is possibly due to either a vertical relationship or domain swapping (Winter et al., 2005). The *pir* orthologue in *P. ovale* spp., herein termed the *oir* multigene family, is highly abundant in copy number, with approximately 1,800–2,100 genes (Table 2), versus 313 genes in *P. vivax* SalI (PlasmoDB-v26 and Carlton et al., 2008) and 200–800 genes, including pseudogenes, in the genomes of rodent malaria parasites (Otto et al., 2014). An unpublished but publicly available version of a clinical isolate genome of *P. vivax* P01 appears to contain a large repertoire of approximately 1,200 *vir* genes ([www.genedb.org/Homepage/PvivaxP01](http://www.genedb.org/Homepage/PvivaxP01)), thus representing the rapid expansion of the *pir* family of variant genes. The *pir* gene family is relatively less expanded in our genome assembly for *P. malariae*, with less than 300 copies (Table 2). A minimum of 42% of PIR sequences in *P. malariae* and in both *P. ovale curtisi* and *P. ovale wallikeri* contain PEXEL-like motifs. It is difficult to determine the precise copy number of the *oir* genes due to the draft nature of the genome assembly, and the majority are relegated to small contigs rather than fully constituted chromosomes. However, our genome



**Fig. 3.** Arrangement of variant gene family arrays in the *Plasmodium ovale wallikeri* nuclear genome. Annotation of genes within four contigs (scaffolds) for *P. ovale wallikeri*. Contig and gene lengths are not drawn to scale. Gene structures are simplified to not show introns. Head to tail arrays are condensed to a single gene icon, and the numbers of genes within the array are given in parentheses. For example, the contig in the top panel contains 36 head to tail *oir* genes. Select gene identifiers are given as letters, and are listed.

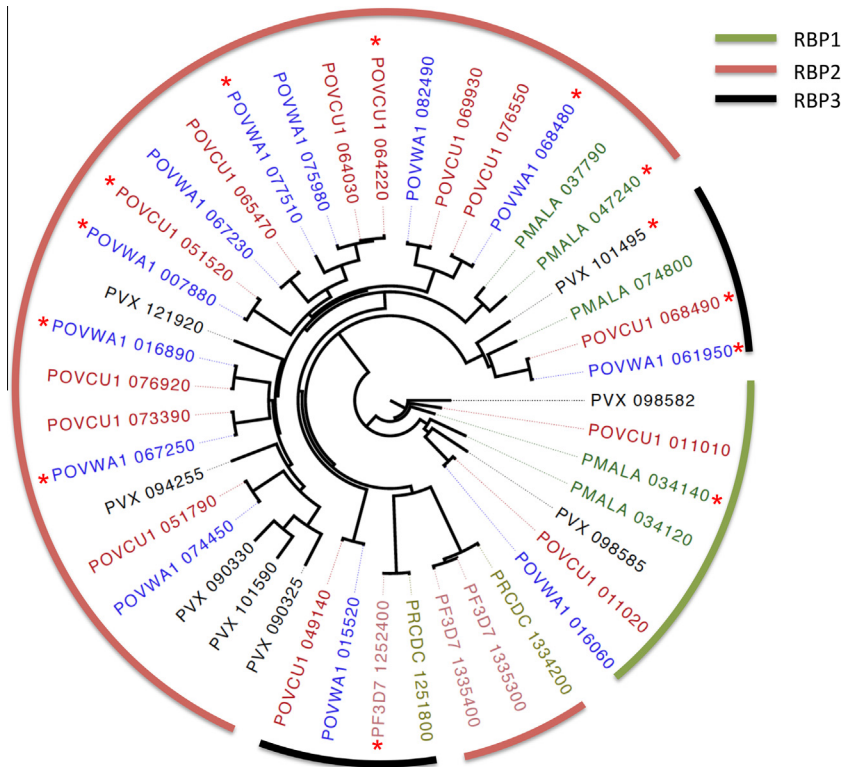
assemblies did produce large contig sequences that harbored *oir* genes, so that some description of their genomic organisation is possible. Large arrays of *oir* genes were found in both *P. ovale curtisi* and *P. ovale wallikeri* genomes, including instances of tandem arrays of up to 39 head-to-tail genes (Fig. 3). The majority of the *surfin* genes are encoded on orphan contigs that could not be placed on a given chromosome, based on synteny with *P. vivax*. These were frequently located within or adjacent to arrays of *oir* genes, supporting the view that a substantial proportion of both gene families are sub-telomeric in position. Proteins orthologous to the SICAvAr and PfEMP1 erythrocyte surface receptors of *P. knowlesi* and *P. falciparum*, respectively, were not found in the *P. ovale* spp. and *P. malariae* predicted proteomes. However, the majority of proteins in the expanded STP1/SURFIN gene family in *P. malariae* and in both *P. ovale* spp. contain a SICA domain. SICAvArs are known to have an evolutionary link with SURFIN/PvSTP1 proteins (Frech and Chen, 2013). Thus the *oir* and *surfin* gene products might be considered the principal candidates to mediate interactions of *P. ovale* parasite-infected erythrocytes and host tissues, and their expansion is likely to have been driven by antibody-mediated immune pressure, as well as evolutionary pressure to recognise multiple host ligands.

The sub-telomeric regions of *P. falciparum* are loosely organised such that the *var* genes, which encode the erythrocyte surface receptors PfEMP1, cluster most telomeric and the highly amplified *rifin* and *stevor* genes, as well as other predicted exported proteins, and localise in the adjacent, more centromeric neighborhoods. Following this progression, towards the centromere are located additional amplified families such as *phist*, *rpb*, *clag*, *lysophospholipase* and *actyl-CoA-synthetase* genes. Examples from the annotation of large contigs suggests that this general theme also holds true for *P. ovale* spp., with the *oir* gene arrays, and interspersed *surfin* genes, replacing the *P. falciparum* *var* genes as most proximal to the telomere (Fig. 3).

### 3.5. Amplified gene families: 2TM superfamily

Annotation of *Plasmodium* genomes identified a superfamily of proteins which contain signal peptides, PEXEL/HT motifs, two predicted TM domains within the C-terminal half of the proteins, and are predicted to be transported to the surface of the infected erythrocytes (Sam-Yellow et al., 2004; Templeton and Deitsch, 2004; Sargeant et al., 2006; Haase et al., 2013). This superfamily, loosely termed 2TM, includes the *P. falciparum* Pfmc2TM, RIFIN and STEVOR proteins (Cheng et al., 1998; Gardner et al., 1998; Kyes et al., 1999), although recent studies suggest that the latter two families encode proteins that possess a single C-terminal TM domain and are likely to be receptors exposed on the erythrocyte surface (Niang et al., 2009, 2014; Bachmann et al., 2015; Goel et al., 2015; Tan et al., 2015). Both *P. ovale* spp. possess a large repertoire of predicted 2TM proteins, with over 90 members per genome including gene fragments. The 2TM proteins were identified by BLAST screens using queries from *P. vivax* and other *Plasmodium* spp., and by annotation of gene neighborhoods containing genes that encode predicted exported proteins such as PHIST domain proteins. Some of the *P. ovale* spp. 2TM proteins have sequence similarity with the Pv-fam-a and Pv-fam-d family proteins of *P. vivax* and non-human primate malaria parasites. Additional *P. ovale* spp. 2TM proteins share similarity with the N-terminal region of *P. falciparum* STEVOR proteins, and over 20 other proteins were identified via BLAST searches. Due to the difficulty in annotation of the overall breadth of 2TM proteins, which relies on structural features in addition to sequence similarity, we have not included enumeration of *Plasmodium* 2TM proteins in Table 2. Examples of *P. ovale* 2TM proteins are listed in Supplementary Table S7.

Notably, predicted 2TM proteins are significantly expanded in *P. malariae*. We have identified more than 550 proteins in *P. malariae* that have 1–6 predicted TM domains with 2TM domain containing



**Fig. 4.** Phylogeny of reticulocyte binding protein (RBP) family of proteins. Phylogenetic relationships of the RBP family of proteins in *Plasmodium ovale wallikeri* (shown in blue), *Plasmodium ovale curtisi* (red), *Plasmodium malariae* (green) and in *Plasmodium vivax* (black) showing expansion of RBP2 members in *P. ovale* genomes. Putative pseudo genes were also included in the analysis by taking the predicted protein sequences. Pseudo genes are denoted with \*. The MrBayes program (Ronquist et al., 2012) was used to infer the phylogenetic tree with 140,000 generations and TreeGraph2 (Stöver and Müller, 2010) was used for visualisation and editing.

proteins being most represented (greater than 75% of the sequences, Table 2, termed Pm-fam-a). More than 40 sequences contain canonical PEXEL/HT erythrocyte targeting motifs and approximately 80 sequences contain PEXEL-like motifs (KxLxD) (Sargeant et al., 2006 and Carlton et al., 2008). More than 350 sequences contain a domain, DUF3671, which is of unknown function (PF12420) (Supplementary Table S4) and which is also present in 33 *P. vivax* and four *P. knowlesi* proteins. *Plasmodium ovale* spp. contain 12–17 copies of these genes. More than 200 DUF3671-containing sequences also have sequence similarity with the pv-fam-b gene family of proteins in *P. vivax* (based on HMM analysis). These proteins do not appear to be related to other 2TM domain containing proteins such as PfMC-2TM, STEVOR or RIFINS or to other known TM domain containing proteins such as PIRs, Tryptophan-rich antigens and PHIST. Thus, these proteins represent a novel *P. malariae*-specific gene, which we term Pm-fam-a.

### 3.6. Amplified families, the tryptophan-rich antigen (*tra*) genes

*Plasmodium ovale wallikeri* and *P. ovale curtisi* share with *P. vivax* and non-human primate malaria parasites an expansion of over 35 copies of *tra* genes encoding the Tryptophan-rich antigen (Burns et al., 1997, 2000; Sutherland et al., 2010). The *tra* genes are typically associated with other expanded gene families such as those described above (Table 2). Both *P. ovale* spp. genomes share extensive orthology among the *tra* loci, suggesting the expansion of this family occurred prior to the separation of the two lineages. *Plasmodium malariae* harbours a lesser level of expansion of the *tra* genes with approximately 30 copies, which could be due to the fragmented and incomplete nature of the genome. *Plasmodium falciparum* possesses four copies of this gene, including a predicted

pseudogene, and six copies are present in rodent malaria parasites; thus the role of Tryptophan-rich antigen proteins in modification of the erythrocyte might be such that *P. ovale* spp. require a greater abundance or functional diversity of the protein, driving expansion of the gene. The tryptophan-rich antigen proteins contain a signal peptide sequence and tryptophan-rich domains. They lack both a canonical PEXEL/HT erythrocyte-targeting motif and a TM domain and so are thought to be exported and function within the erythrocyte cytoplasm. It was recently suggested that the Tryptophan-rich proteins interact with erythrocyte membranes B and three (Alam et al., 2015).

### 3.7. Amplified families: the reticulocyte binding-like (RBL) and erythrocyte binding-like (EBL) proteins

Studies on the invasion of reticulocytes and normocytes by *P. knowlesi*, *P. vivax* and *P. falciparum* led to the discovery of merozoite surface receptors, including the Duffy binding and reticulocyte binding-like proteins (RBL, also called RBP) (reviewed in Gaur et al., 2004; Culleton and Kaneko, 2010; Li and Han, 2012; Gunalan et al., 2013). *Plasmodium falciparum* possesses five members of the RBL family, whereas *P. vivax* encodes 10 members including pseudogenes. In *P. ovale* spp. the *rbl* genes are expanded to 12–13 copies, including several putative pseudogenes (please see Table 2 for details), and were found within or adjacent to arrays of *oir* genes (Fig. 3). In contrast, *P. knowlesi* appears to have only two members of the family, thus highlighting the range in size of the RBL protein repertoires across *Plasmodium* spp. Similar to primate malaria parasites and several *P. vivax* isolates, the genomes of *P. ovale* spp. and *P. malariae* contain an expansion (two copies) of the erythrocyte binding-like proteins (EBL; also called Duffy



binding-like, DBL). *Plasmodium vivax* Sal I possesses a single DBL gene, but DBP is known to be duplicated in *P. vivax* isolates found in other geographic regions such as East Africa and non-African *P. vivax* endemic countries (Menard et al., 2013). BLAST analyses as well as construction of gene trees suggest that the *P. ovale* expansion derives from RBP2 (Fig. 4). In our datasets, *P. ovale curtisi* and *P. ovale wallikeri* have seven and four functionally intact RBP2 genes, respectively, and the rest of the expanded RBP2 genes appear to have been pseudogenised by acquiring in-frame stop codons with the coding sequences. In contrast, *P. malariae* maintains three functional RBP genes (one each of RBP1, 2 and 3) and has two RBP pseudogenes (RBP1, RBP2) (Supplementary Table S4). The species-specific expansion and diversification of the RBL family of genes may be related to the phenotypic diversity that these parasites show with respect to their red cell preferences. While *P. ovale* spp. are restricted to reticulocytes, *P. malariae* invades normocytes (Garnham, 1966; Gunalan et al., 2013).

### 3.8. Amplified families: the ookinete surface and gametocyte stage proteins

The *Plasmodium* ookinete surface protein P28 is encoded by four genes whose expansion appears to be unique to *P. ovale* spp. (Table 2). The gene was first isolated from *Plasmodium gallinaceum* (Duffy et al., 1993), and in all *Plasmodium* parasites is found as a single copy, adjacent to the gene encoding the ookinete protein P25. The two ookinete surface proteins share a structure of signal peptide, multiple EGF domains, and a predicted GPI anchor signal sequence at the C-terminus. In *P. ovale* spp. the *p28* gene was reported to be duplicated (Tachibana et al., 2001), and now complete genome information for *P. ovale wallikeri* reveals two additional copies of the gene. One *p28* copy (termed *p28-1*; Tachibana et al., 2001) is adjacent to *p25*, with conserved synteny with neighboring genes in the locus across *Plasmodium*, and is presumed to be the ancestral *p28*, related to *p25* via a gene duplication event. The other three genes share greater amino acid similarity (80–97%) and are divergent from the ancestral P28 (<50%). Of the expanded copies, one *p28* gene is within an orphan contig; and two are within contigs that contain *oir* genes, one of which (termed *p28-2*; Tachibana et al., 2001) is within a large array of *oir* genes. *Plasmodium ovale curtisi* also has an expansion of *p28* genes. P28 is not expanded in our assembly for *P. malariae*.

### 3.9. Gene expansion encoding the P27/25 gamete protein in *P. malariae*

The P27/25 protein is a highly abundant cytoplasmic protein expressed during early gametogenesis in *P. falciparum* (Carter et al., 1989). Disruption of the P27/25 gene affects membrane integrity during gametocyte maturation and results in the appearance of large intracellular vacuoles and membrane abnormalities (Olivieri et al., 2009). The protein lacks a signal peptide sequence or predicted TM domain, and possesses proline motifs (PxxP) at the N- and C-terminal regions that are proposed to interact with RNA or other proteins (Sharma et al., 2003; Camarda et al., 2010). P27/25 has been identified in *P. falciparum*, *P. knowlesi* and *P. reichenowi*, although it might share a domain with other conserved *Plasmodium* proteins which possess predicted signal peptide sequence but lack PxxP motifs (eg. PVX\_003560 in *P. vivax*). *Plasmodium malariae* harbours an expansion of 16 P27/25 genes and is thus the second *Plasmodium* sp. identified to possess P27/25 proteins, and uniquely, an expansion thereof (Table 2). Many, but not all, of the *P. malariae* versions possess one or multiple PxxP motifs and several encode proteins with possible signal peptide sequences (amino acid alignment shown in Supplementary Fig. S1); thus the

gene models for all examples should be validated to confirm the proposed protein structures.

### 3.10. A novel *P. ovale*-specific amplification: the KELT proteins

*Plasmodium ovale wallikeri* and *P. ovale curtisi* possess a tandem head to tail array expansion of eight genes encoding an uncharacterised protein, herein named KELT for a C-terminal motif. The pairwise amino acid similarities range from 40 to 65%, indicating a selective pressure mediating divergence of members of the family (amino acid alignment shown in Supplementary Fig. S2). The expansion appears to be specific to *P. ovale* spp., as the gene is present as a single copy in all other primate *Plasmodium* genomes elucidated to date, and is absent from the genomes of the rodent malaria parasites (Table 2). Synteny is conserved with *P. falciparum*, which maps to a sub-telomeric region. The *P. ovale* spp. and *P. vivax* versions encode signal peptide sequences, which are lacking in the single exon gene model of the *P. falciparum* version (PF3D7\_1475900). However, a putative signal peptide is encoded in an alternative single exon gene model for the *P. falciparum* locus, in which translation is initiated from an in-frame upstream methionine; and this 13 amino acid long extension identifies other KELT N-terminal regions in BLAST queries (Supplementary Fig. S2). The KELT protein is not expanded in our assembly for *P. malariae* (Table 2).

### 3.11. Synteny and antigenic divergence between the *P. ovale* spp

We compared the conservation of gene order between the *P. ovale* spp. and with *P. vivax*. There were no apparent large-scale rearrangements between the contiguated DNAs of *P. ovale wallikeri* and *P. ovale curtisi*, based on comparative ordering of contigs on to the assembled *P. vivax* Sal1 genome (Supplementary Fig. S3). However, due to fragmentation of the genome assemblies, our observation is restricted to the set of contigs (~22 Mb of the total contiguated bp in the assembly) that we could align with high confidence. However, micro-synteny breaks between *P. ovale* spp. were identified as described below. Due to a higher level of fragmentation of the *P. malariae* genome, we did not attempt to study synteny with other *Plasmodium* spp. in this study.

The largest contig in the *P. ovale wallikeri* assembly encompasses a predicted sub-telomeric region in which 87 of the 93 predicted CDSs are encoded by *oir* genes. This contig is conserved in the second isolate of *P. ovale wallikeri*, with 84 syntenic genes, based upon location and similarity via BLAST analysis. Synteny is disrupted by an insertion which appears to have arisen by duplication of an adjacent segment (Supplementary Table S5). The syntenic *oir* genes range between 100% and 32% amino acid identity, and it thus appears that selection pressures underpinning amino acid diversity act differentially on neighbouring genes. The fragmented state of the *P. ovale curtisi* assembly precluded description of synteny within this region, although BLAST analyses using as queries the 87 *P. ovale wallikeri* *oir* gene products from Supplementary Table S6 suggest orthologous relationships exist in *P. ovale curtisi*, but with amino acid identities no greater than 90% (results from a subset of *oir* genes shown in Supplementary Table S6). The fact that identical *oir* genes were readily found in our two *P. ovale wallikeri* isolates, versus no similarities greater than 90% between *oir* genes in *P. ovale curtisi* and *P. ovale wallikeri*, highlights the genetic difference between the two *ovale* malaria parasite species. Similarly, SURFIN proteins were frequently identical between the two *P. ovale wallikeri* isolates, but were more divergent between *P. ovale curtisi* and *P. ovale wallikeri* (Supplementary Table S6), with the exception of one identified SURFIN (POVWA1\_083840) that was invariant between the two species. The SURFIN repertoires between the two parasites differed such that divergent *P. ovale*

*wallikeri* SURFIN queries frequently retrieved the same best hits as in *P. ovale curtisi* (Supplementary Table S6).

The intra- and inter-species diversity of other parasite and erythrocyte surface antigens was also examined. The RBL and EBL proteins are identical between the two *P. ovale wallikeri* isolates, with 100% amino acid conservation, whereas there was no more than 92% conservation inter-species in comparison with predicted orthologs in *P. ovale curtisi* (Supplementary Table S6). Similarly, the AMA1 and P25 proteins are identical between the two *P. ovale wallikeri* isolates, whereas they are 95% conserved inter-species. MSP1 and CSP1 show more antigenic divergence between the two *P. ovale wallikeri* isolates, 95% and 93% respectively, while again retaining greater divergence across species with *P. ovale curtisi*, 83% and 68% respectively. The erythrocyte cytoplasm-localised PHIST domain and Tryptophan-rich protein families also follow this trend, with typically 100% identity between intra-species isolates but less than 94% amino acid similarity between the two *P. ovale* spp. (data not shown). For the above antigens, plus OIR and SURFIN proteins, the two *P. ovale curtisi* isolates were also more similar to each other than to *P. ovale wallikeri* (data not shown). Thus the intra- versus inter-species diversity of these antigens supports the genetic isolation and distance between the two species.

### 3.12. Whole genome variation analysis

We have identified 450,033 fixed nucleotide differences between both Poc1 and Poc2 isolates with Pow1 by analyzing common SNPs derived from Poc1 and Poc2 Illumina reads mapping to the Pow1 reference. Out of these SNPs, we found 219,947 SNPs in the coding regions, nearly half of which (109,727) were non-synonymous SNPs, which results in approximately one nucleotide difference for every 78 bases between *P. ovale curtisi* and *P. ovale wallikeri*. We repeated this analysis by using the Pow2 genome as a reference and detected an equivalent number (453,165) of SNPs. Similarly, we observed 26,361 common small insertions and 30,379 common deletion events in both *P. ovale curtisi* isolates compared with *P. ovale wallikeri*. The detailed list of SNPs, indels statistics and gene annotations are provided in Supplementary Tables S8–S10.

### 3.13. Concluding comments

The repertoire of expanded proteins in *P. ovale curtisi* and *P. ovale wallikeri* confirm their phylogenetic relationship to *P. vivax* and *P. knowlesi*, and distance from *P. falciparum*. The genomes of *P. ovale* spp. possess a greatly expanded *pir* gene family, herein referred to as *oirs*, but lack families homologous to either the PfEMP1 erythrocyte surface receptors of *P. falciparum*, or the SICA-var proteins of *P. knowlesi*. Moreover, *P. ovale curtisi* and *P. ovale wallikeri* utilise Duffy binding protein homologues for erythrocyte invasion, rather than ligands related to the EBL proteins of *P. falciparum*. Similar to *P. vivax*, both *ovale* parasite genomes harbour an expansion of RBP2 proteins. Unique to both *P. ovale* and *P. malariae*, is an expansion of *surfin* genes, thus suggesting an important role, as yet unknown, for this parasite-encoded predicted erythrocyte surface receptor. Species status for *P. ovale wallikeri* and *P. ovale curtisi* is supported by their separation in the phylogenetic trees, and the observation that SURFIN, OIR and other parasite and erythrocyte surface proteins were similar in the two *P. ovale wallikeri* isolates, as well as the two *P. ovale curtisi* isolates, but uniformly more divergent in comparisons of predicted orthologs between the two species. Availability of the *P. malariae* draft genome also provides us with a comparative tool to study the expansion of variant gene families across malaria parasite species. The presence of the gene encoding the abundant gametocyte P25/27 protein in *P. malariae* unites it with *P. falciparum*, *P. knowlesi* and *P. reichenowi*,

whereas the expansion is unique to this species. A unique expansion of a highly diverged TM protein family (Pm-fam-a) with over 550 members delineates the *P. malariae* genome from the rest of the malaria parasite genomes available to date. These may play a major role in its interactions with the host during infection processes.

The three human malaria parasite genomes described here provide new resources to study human malaria parasite biology and underpin future comparative genomics studies covering all human-infecting malaria species.

### Acknowledgment

TJT was supported by a visiting professorship to the Institute of Tropical Medicine, Nagasaki University, Japan. JC was supported by grants from the National Natural Science Foundation of China (No. 81271870) and the National Natural Science Foundation of Jiangsu Province, China (No. BK20150001). AP is supported by faculty baseline funding from the King Abdullah University of Science and Technology (KAUST), Saudi Arabia and the Global Institute for Collaborative Research and Education (GI-CoRE) at the Research Center for Zoonosis Control, Hokkaido University, Japan. RC is supported by Japanese Society of Parasitology (JSPS), Japan Grant-in-Aid for Scientific Research Nos. 24255009, 25870525 and 16K21233. MCO, EDB, TGC and CJS are supported by UK Medical Research Council Project Grant MR/L008661/1 to CJS. The authors thank the staff of the Bioscience Core Laboratory in KAUST for sequencing the DNA libraries. Osamu Kaneko is thanked for discussions and critical reading of the manuscript.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ijpara.2016.05.009>.

### References

- Alam, M.S., Choudhary, V., Zeeshan, M., Tyagi, R.K., Rathore, S., Sharma, Y.D., 2015. Interaction of *Plasmodium vivax* tryptophan-rich antigen PvTRAg38 with Band 3 on human erythrocyte surface facilitates parasite growth. *J. Biol. Chem.* 290, 20257–20272.
- Alexandre, J.S., Yahata, K., Kawai, S., Torii, M., Kaneko, O., 2011. PEXEL-independent trafficking of *Plasmodium falciparum* SURFIN4.2 to the parasite-infected red blood cell and Maurer's clefts. *Parasitol. Int.* 60, 313–320.
- Arisue, N., Hashimoto, T., Mitsui, H., Palacpac, N.M.Q., Kaneko, A., Kawai, S., Hasegawa, M., Tanabe, K., Horii, T., 2012. The *Plasmodium* apicomplast genome: Conserved structure and close relationship of *P. ovale* to rodent malaria parasites. *Mol. Biol. Evol.* 29, 2095–2099.
- Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., Berriman, M., 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25, 1968–1969.
- Assefa, S.A., Preston, M.D., Campino, S., Ocholla, H., Sutherland, C.J., Clark, T.G., 2014. EstMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinformatics* 30, 1292–1294.
- Bachmann, A., Scholz, J.A.M., Janßen, M., Klinkert, M.Q., Tannich, E., Bruchhaus, I., Petter, M., 2015. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in *Plasmodium falciparum*-infected erythrocytes. *Malar. J.* 14, 274.
- Bauffe, F., Desplans, J., Fraissier, C., Parzy, D., 2012. Real-time PCR assay for discrimination of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* in the Ivory Coast and in the Comoros Islands. *Malar. J.* 11, 307.
- Boddey, J.A., Carvalho, T.G., Hodder, A.N., Sargeant, T.J., Sleebs, B.E., Marapana, D., Lopaticki, S., Nebl, T., Cowman, A.F., 2013. Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome. *Traffic* 14, 532–550.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Burns, J.M., Dunn, P.D., Russo, D.M., 1997. Protective immunity against *Plasmodium yoelii* malaria induced by immunization with particulate blood-stage antigens. *Infect. Immun.* 65, 3138–3145.

- Burns, J.M., Adeeku, E.K., Belk, C.C., Dunn, P.D., 2000. An unusual tryptophan-rich domain characterizes two secreted antigens of *Plasmodium yoelii*-infected erythrocytes. *Mol. Biochem. Parasitol.* 110, 111–121.
- Camarda, G., Bertuccini, L., Singh, S.K., Salzano, A.M., Lanfancotti, A., Olivieri, A., Scaloni, A., Sharma, A., Alano, P., 2010. Regulated oligomerization and molecular interactions of the early gametocyte protein Pfg27 in *Plasmodium falciparum* sexual differentiation. *Int. J. Parasitol.* 40, 663–673.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973.
- Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S.V., Merino, E.F., Amedeo, P., Cheng, Q., Coulson, R.M., Crabb, B.S., Del Portillo, H.A., Essien, K., Feldblyum, T.V., Fernandez-Becerra, C., Gilson, P.R., Gueye, A.H., Guo, X., Kang'a, S., Kooij, T.W., Korsinczyk, M., Meyer, E.V., Nene, V., Paulsen, I., White, O., Ralph, S.A., Ren, Q., Sargeant, T.J., Salzberg, S.L., Stockert, C.J., Sullivan, S.A., Yamamoto, M.M., Hoffman, S.L., Wortman, J.R., Gardner, M.J., Galinski, M.R., Barnwell, J.W., Fraser-Liggett, C.M., 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455, 757–763.
- Carter, R., Graves, P.M., Creasey, A., Byrne, K., Read, D., Alano, P., Fenton, B., 1989. *Plasmodium falciparum*: an abundant stage-specific protein expressed during early gametocyte development. *Exp. Parasitol.* 69, 140–149.
- Carver, T., Harris, S.R., Otto, T.D., Berriman, M., Parkhill, J., McQuillan, J.A., 2013. BamView: visualizing and interpretation of next-generation sequencing read alignments. *Briefings Bioinf.* 14, 203–212.
- Chaturvedi, N., Bhandari, S., Bharti, P.K., Basak, S.K., Singh, M.P., Singh, N., 2015. Sympatric distribution of *Plasmodium ovale curtisi* and *P. ovale wallikeri* in India: implication for the diagnosis of malaria and its control. *Trans. R. Soc. Trop. Med. Hyg.* 109, 352–354.
- Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waite, G., Lanzer, M., Saul, A., 1998. *Stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol. Biochem. Parasitol.* 97, 161–176.
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92.
- Coil, D., Jospin, G., Darling, A.E., 2015. A5-misec: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 31, 587–589.
- Collins, W.E., Schwartz, I.K., Skinner, J.C., Broderson, J.R., 1984. Studies on the Uganda I/CDC strain of *Plasmodium malariae* in Bolivian *Aotus* monkeys and various anophelines. *J. Parasitol.* 70, 677–681.
- Collins, W.E., Pappaioanou, M., McClure, H.M., Swenson, R.B., Strobert, E., Skinner, J. C., Filipiński, V., Stanfill, P.S., Colliins, F.H., Campbell, C.C., 1987. Infection of chimpanzees with Nigerian I/CDC strain of *Plasmodium ovale*. *Am. J. Trop. Med. Hyg.* 37, 455–459.
- Collins, W.E., McClure, H.M., Strobert, E., Filipiński, V., Skinner, J.C., Stanfill, P.S., Richardson, B.B., Morris, C., 1990. Infection of chimpanzees with the Uganda I/CDC strain of *Plasmodium malariae*. *Am. J. Trop. Med. Hyg.* 42, 99–103.
- Culleton, R., Kaneko, O., 2010. Erythrocyte binding ligands in malaria parasites: intracellular trafficking and parasite virulence. *Acta Trop.* 114, 131–137.
- Cunningham, D., Lawton, J., Jarra, W., Preiser, P., Langhorne, J., 2010. The *pir* multigene family of *Plasmodium*: Antigenic variation and beyond. *Mol. Biochem. Parasitol.* 170, 65–73.
- del Portillo, H.A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C.P., Schneider, N.K., Villalobos, J.M., Rajandream, M.A., Harris, D., Pereira da Silva, L.H., Barrell, B., Lanzer, M., 2001a. A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* 410, 839–842.
- del Portillo, H.A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Peruss, M., Sanchez, C.P., Schneider, N.K., Villalobos, J.M., Rajandream, M.A., Harris, D., Pereira da Silva, L.H., Barrell, B., Lanzer, M., 2001b. A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* 410, 839–842.
- Dinko, B., Oguike, M.C., Larbi, J.B., Bousema, J.T., Sutherland, C.J., 2013. Persistent detection of *Plasmodium falciparum*, *P. malariae*, *P. ovale curtisi* and *P. ovale wallikeri* after ACT treatment of asymptomatic Ghanaian school-children. *Int. J. Parasitol.* 43, 45–50.
- Duffy, P.E., Pimenta, P., Kaslow, D.C., 1993. Pgs28 belongs to a family of epidermal growth factor-like antigens that are targets of malaria transmission-blocking antibodies. *J. Exp. Med.* 177, 505–510.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Fançon, C., Gamboa, D., Sebastião, Y., Hallett, R., Sutherland, C., Sousa-Figueiredo, J. C., Nery, S.V., 2012. Various *pfcr* and *pfmdr1* genotypes of *Plasmodium falciparum* co-circulate with *P. malariae*, *P. ovale spp* and *P. vivax* in northern Angola. *Antimicrob. Agents Chemother.* 56, 5271–5277.
- Faye, F.B., Konaté, L., Rogier, C., Trape, J.F., 1998. *Plasmodium ovale* in a highly malaria endemic area of Senegal. *Trans. R. Soc. Trop. Med. Hyg.* 92, 522–525.
- Frech, C., Chen, N., 2013. Variant surface antigens of malaria parasites: functional and evolutionary insights from comparative gene family classification and analysis. *BMC Genomics* 14, 427.
- Fuehrer, H.P., Habler, V.E., Fally, M.A., Harl, J., Starzengruber, P., Swoboda, P., Bloesch, I., Khan, W.A., Noedl, H., 2012. *Plasmodium ovale* in Bangladesh: Genetic diversity and the first known evidence of the sympatric distribution of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* in southern Asia. *Int. J. Parasitol.* 42, 693–699.
- Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Perteau, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C., Hoffman, S.L., 1998. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* 282, 1126–1132.
- Garnham, P.C.C., 1966. *Malaria Parasites and other Hemosporidia*. Blackwell Scientific Publications, Oxford, UK.
- Gaur, D., Mayer, D.C.G., Miller, L.H., 2004. Parasite-host receptor interactions during invasion of erythrocytes by *Plasmodium* merozoites. *Int. J. Parasitol.* 34, 1413–1429.
- Goel, S., Palmkvist, M., Moll, K., Joannin, N., Lara, P., Akhouri, R.R., Moradi, N., Öjemalm, K., Westman, M., Angeletti, D., Kjellen, H., Lehtiö, J., Blixt, O., Ideström, L., Gahmberg, C.G., Storry, J.R., Hult, A.K., Olsson, M.L., von Heijne, G., Nilsson, I., Wahlgren, M., 2015. RIFINs are adhesins implicated in severe *Plasmodium falciparum* malaria. *Nat. Med.* 21, 314–317.
- Grassi, B., Feletti, R., 1892. Contribuzione allo studio dei parassiti malarici. *Atti Accad. Gioenia Sci. nat. Catania, Memoir* 5, 1–80.
- Gunalan, K., Gao, X., Yap, S.S., Huang, X., Preiser, P.R., 2013. The role of the reticulocyte-binding-like protein homologues of *Plasmodium* in erythrocyte sensing and invasion. *Cell. Microbiol.* 15, 35–44.
- Haase, S., Hanssen, E., Matthews, K., Kalanon, M., de Koning-Ward, T.F., 2013. The exported protein PbCP2 localises to cleft-like structures in the rodent malaria parasite *Plasmodium berghei*. *PLoS One* 8, e61482.
- Hawthorne, P.L., Trenholme, K.R., Skinner-Adams, T.S., Spielman, T., Fischer, K., Dixon, M.W., Ortega, M.R., Anderson, K.L., Kemp, D.J., Gardiner, D.L., 2004. A novel *Plasmodium falciparum* ring stage protein, REX, is located in Maurer's clefts. *Mol. Biochem. Parasitol.* 146, 181–189.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645.
- Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12.
- Kyes, S.A., Rowe, J.A., Kriek, N., Newbold, C.I., 1999. Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9333–9338.
- Lau, Y.L., Lee, W.C., Tan, L.H., Kamarulzaman, A., Syed Omar, S.F., Fong, M.Y., Cheong, F.W., Mahmud, R., 2013. Acute respiratory distress syndrome and acute renal failure from *Plasmodium ovale* infection with fatal outcome. *Malar. J.* 12, 389.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, J., Han, E.T., 2012. Dissection of the *Plasmodium vivax* reticulocyte binding-like proteins (PvRBP). *Biochem. Biophys. Res. Commun.* 426, 1–6.
- Li, L., Stoehcher, C.J., Roos, D.S., 2003. OrthoMCL: identification of orthologous groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189.
- Logan-Klumpler, F.J., De Silva, N., Boehme, U., Rogers, M.B., Velarde, G., McQuillan, J. A., Carver, T., Aslett, M., Olsen, C., Subramanian, S., Phan, I., Farris, C., Mitra, S., Ramasamy, G., Wang, H., Tivey, A., Jackson, A., Houston, R., Parkhill, J., Holden, M., Harb, O.S., Brunk, B.P., Myler, P.J., Roos, D., Carrington, M., Smith, D.F., Hertz-Fowler, C., Berriman, M., 2012. GeneDB – an annotation database for pathogens. *Nucleic Acids Res.* 40, D98–D108.
- Lopez, F.J., Bernabeu, M., Fernandez-Becerra, C., del Portillo, H.A., 2013. A new computational approach redefines the subtelomeric *vir* superfamily of *Plasmodium vivax*. *BMC Genomics* 14, 8.
- Lysenko, A.J., Beljaev, A.E., 1969. An analysis of the geographical distribution of *Plasmodium ovale*. *Bull. World Health Org.* 40, 383–394.
- May, J., Mockenhaupt, F.P., Ademowo, O.G., Falusi, A.G., Olumese, P.E., Bienze, U., Meyer, C.G., 1999. High rate of mixed and subpatent malarial infections in southwest Nigeria. *Am. J. Trop. Med. Hyg.* 61, 339–343.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Menard, D., Chan, E.R., Benedet, C., Ratsimbaoa, A., Kim, S., Chim, P., Do, C., Witkowski, B., Durand, R., Thellier, M., Severini, C., Legrand, E., Musset, L., Nour, B.Y., Mercereau-Pujalon, O., Serre, D., Zimmerman, P.A., 2013. Whole genome sequencing of field isolates reveals a common duplication of the Duffy binding protein gene in Malagasy *Plasmodium vivax* strains. *PLoS Negl. Trop. Dis.* 7, e2489.
- Mueller, I., Zimmerman, P.A., Reeder, J.C., 2007. *Plasmodium malariae* and *Plasmodium ovale* – the “bashful” malaria parasites. *Trends Parasitol.* 23, 278–283.
- Niang, M., Yan Yam, X., Preiser, P.R., 2009. The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog.* 5, e1000307.
- Niang, M., Bei, A.K., Madnani, K.G., Pelly, S., Dankwa, S., Kanjee, U., Gunalan, K., Amaladoss, A., Yeo, K.P., Bob, N.S., Malleret, B., Duraisingham, M.T., Preiser, P.R., 2014. STEVOR is a *Plasmodium falciparum* erythrocyte binding protein that mediates merozoite invasion and resetting. *Cell Host Microbe* 16, 81–93.
- Oguike, M.C., Betson, M., Burke, M., Nolder, D., Stothard, R., Kleinschmidt, I., Proietti, C., Bousema, T., Ndounga, M., Tanabe, K., Ntege, E., Culleton, R., Sutherland, C.J., 2011. *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* circulate simultaneously in African communities. *Int. J. Parasitol.* 41, 677–683.
- Oguike, M.C., Sutherland, C.J., 2015. Dimorphism in genes encoding sexual-stage proteins of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. *Int. J. Parasitol.* 45, 449–454.
- Olivieri, A., Camarda, G., Bertuccini, L., van de Vegte-Bolmer, M., Luty, A.J.F., Sauerwein, R., Alano, P., 2009. The *Plasmodium falciparum* protein Pfg27 is

- dispensable for gametocyte and gamete production, but contributes to cell integrity during gametocytogenesis. *Mol. Microbiol.* 73, 180–193.
- Otto, T.D., Böhme, U., Jackson, A.P., Hunt, M., Franke-Fayard, B., Hoeijmakers, W.A.M., Religa, A.A., Robertson, L., Sanders, M., Ogun, S.A., Cunningham, D., Erhart, A., Billker, O., Khan, S.M., Stunnenberg, H.G., Langhorne, J., Holder, A.A., Waters, A.P., Newbold, C.I., Pain, A., Berriman, M., Janse, C.J., 2014. A comprehensive evaluation of rodent malaria genome and gene expression. *BMC Biol.* 12, 86.
- Putaporntip, C., Hughes, A.L., Jongwutiwes, S., 2013. Low level of sequence diversity at merozoite surface protein-1 locus of *Plasmodium ovale curtisi* and *P. ovale wallikeri* from Thai isolates. *PLoS One* 8 e58962.
- Rojo-Marcos, G., Cuadros-González, J., Mesa-Latorre, J.M., Culebras-López, A.M., de Pablo-Sánchez, R., 2008. Acute respiratory distress syndrome in a case of *Plasmodium ovale* malaria. *Am. J. Trop. Med. Hyg.* 79, 391–393.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945.
- Sam-Yellow, T.Y., Florens, L., Johnson, J.R., Wang, T., Drazba, J.A., Le Roch, K.G., Zhou, Y., Batalov, S., Carucci, D.J., Winzeler, E.A., Yates, J.R., 2004. A *Plasmodium* gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling. *Genome Res.* 14, 1052–1059.
- Schaer, J., Perkins, S.L., Decher, J., Leendertz, F.H., Fahr, J., Weber, N., Matuschewski, K., 2013. High diversity of West African bat malaria parasites and a tight link with rodent *Plasmodium* taxa. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17415–17419.
- Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., Cowman, A.F., 2006. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* 7, R12.
- Sharma, A., Sharma, I., Kogkasuriyachai, D., Kumar, N., 2003. Structure of a gametocyte protein essential for sexual development in *Plasmodium falciparum*. *Nat. Struct. Biol.* 10, 197–203.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212.
- Stanke, M., Schöffmann, O., Morgenstern, B., Waack, S., 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7, 62.
- Stephens, J.W.W., 1922. A new malaria parasite of man. *Ann. Trop. Med. Parasitol.* 16, 383–388.
- Stothard, P., 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28 (1102), 1104.
- Stöver, B.C., Müller, K.F., 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 11, 7.
- Strydom, K.A., Ismail, F., Frean, J., 2014. *Plasmodium ovale*: a case of not-so-benign tertian malaria. *Malar. J.* 13, 85.
- Sutherland, C.J., Tanamsing, N., Nolder, D., Oguike, M., Jennison, C., Pukrittayakamee, S., Dolecek, C., Hien, T.T., Dorosário, V.E., Arez, A.P., Pinto, J., Michon, P., Escalante, A.A., Nosten, F., Burke, M., Lee, R., Blaze, M., Otto, T.D., Barnwell, J.W., Pain, A., Williams, J., White, N.J., Day, N.P.J., Snounou, G., Lockhart, P.J., Chiodini, P.L., Imwong, M., Polley, S.D., 2010. Two nonrecombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. *J. Infect. Dis.* 201, 1544–1550.
- Swain, M.T., Tsai, I.J., Assefa, S.A., Newbold, C., Berriman, M., Otto, T.D., 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat. Protoc.* 7, 1260–1284.
- Tachibana, M., Tsuboi, T., Templeton, T.J., Kaneko, O., Torii, M., 2001. Presence of three distinct ookinete protein genes, Pos25, Po28-1, and Pos28-2, in *Plasmodium ovale*. *Mol. Biochem. Parasitol.* 113, 341–344.
- Tachibana, M., Tsuboi, T., Kaneko, O., Khuntirat, B., Torii, M., 2002. Two types of *Plasmodium ovale* defined by SSU rRNA have distinct sequences for ookinete surface proteins. *Mol. Biochem. Parasitol.* 122, 223–226.
- Tachibana, S., Sullivan, S.A., Kawai, S., Nakamura, S., Kim, H.R., Goto, N., Arisue, N., Palacpac, N.M.Q., Honma, H., Yagi, M., Tougan, T., Kataikai, Y., Kaneko, O., Mita, T., Kita, K., Yasutomi, Y., Sutton, P.L., Shakhbatyan, R., Horii, T., Yasunaga, T., Barnwell, J.W., Escalante, A.A., Carlton, J.M., Tanabe, K., 2012. *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat. Genet.* 44, 1051–1057.
- Tao, Z.Y., Xia, H., Cao, J., Gao, Q., 2011. Development and evaluation of a prototype non-woven fabric filter for purification of malaria-infected blood. *Malar. J.* 10, 251.
- Tan, J., Pieper, K., Piccoli, L., Abdi, A., Foglierini, M., Geiger, R., Tully, C.M., Jarrossay, D., Ndungu, F.M., Wambua, J., Bejon, P., Fregni, C.S., Fernandez-Rodriguez, B., Barbieri, S., Bianchi, S., Marsh, K., Thathy, V., Corti, D., Sallusto, F., Bull, P., Lanzavecchia, A., 2015. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. *Nature* 529, 105–109.
- Templeton, T.J., Deitsch, K.W., 2004. Targeting malaria parasite proteins to the erythrocyte. *Trends Parasitol.* 21, 299–402.
- Tobian, A.A., Mehlotra, R.K., Malhotra, I., Wamachi, A., Mungai, P., Koech, D., Ouma, J., Zimmerman, P., King, C.L., 2000. Frequent umbilical cord-blood and maternal-blood infections with *Plasmodium falciparum*, *P. malariae*, and *P. ovale* in Kenya. *J. Infect. Dis.* 182, 558–563.
- Tomar, L.R., Giri, S., Baudhdh, N.K., Jhamb, R., 2015. Complicated malaria: a rare presentation of *Plasmodium ovale*. *Trop. Doct.* 45, 140–142.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9 e112963.
- Winn, T.T., Jalloh, A., Tantular, I.S., Tsuboi, T., Ferreira, M.U., Kimura, M., Kawamoto, F., 2004. Molecular analysis of *Plasmodium ovale* variants. *Emerging Infect. Dis.* 10, 1235–1240.
- Winter, G., Kawai, S., Haeggström, M., Kaneko, O., von Euler, A., Kawazu, S., Palm, D., Fernandez, V., Wahlgren, M., 2005. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *J. Exp. Med.* 201, 1853–1863.
- Zhu, X., Yahata, K., Alexandre, J.S., Tsuboi, T., Kaneko, O., 2013. The N-terminal segment of *Plasmodium falciparum* SURFIN4.1 is required for its trafficking to the red blood cell cytosol through the endoplasmic reticulum. *Parasitol. Int.* 62, 215–219.