CrossMark

# Multiplicity and molecular epidemiology of *Plasmodium vivax* and *Plasmodium falciparum* infections in East Africa

Daibin Zhong[1*] , Eugenia Lo[2], Xiaoming Wang[1], Delenasaw Yewhalaw[3,4], Guofa Zhou[1], Harrysone E. Atieli[5], Andrew Githeko[5], Elizabeth Hemming-Schroeder[1], Ming-Chieh Lee[1], Yaw Afrane[6] and Guiyun Yan[1*]

## Abstract

**Background:** Parasite genetic diversity and multiplicity of infection (MOI) affect clinical outcomes, response to drug treatment and naturally-acquired or vaccine-induced immunity. Traditional methods often underestimate the frequency and diversity of multiclonal infections due to technical sensitivity and specificity. Next-generation sequencing techniques provide a novel opportunity to study complexity of parasite populations and molecular epidemiology.

**Methods:** Symptomatic and asymptomatic *Plasmodium vivax* samples were collected from health centres/hospitals and schools, respectively, from 2011 to 2015 in Ethiopia. Similarly, both symptomatic and asymptomatic *Plasmodium falciparum* samples were collected, respectively, from hospitals and schools in 2005 and 2015 in Kenya. Finger-pricked blood samples were collected and dried on filter paper. Long amplicon (> 400 bp) deep sequencing of merozoite surface protein 1 (*msp1*) gene was conducted to determine multiplicity and molecular epidemiology of *P. vivax* and *P. falciparum* infections. The results were compared with those based on short amplicon (117 bp) deep sequencing.

**Results:** A total of 139 *P. vivax* and 222 *P. falciparum* samples were pyro-sequenced for *pvmsp1* and *pfmsp1*, yielding a total of 21 *P. vivax* and 99 *P. falciparum* predominant haplotypes. The average MOI for *P. vivax* and *P. falciparum* were 2.16 and 2.68, respectively, which were significantly higher than that of microsatellite markers and short amplicon (117 bp) deep sequencing. Multiclonal infections were detected in 62.2% of the samples for *P. vivax* and 74.8% of the samples for *P. falciparum*. Four out of the five subjects with recurrent *P. vivax* malaria were found to be a relapse 44–65 days after clearance of parasites. No difference was observed in MOI among *P. vivax* patients of different symptoms, ages and genders. Similar patterns were also observed in *P. falciparum* except for one study site in Kenyan lowland areas with significantly higher MOI.

**Conclusions:** The study used a novel method to evaluate *Plasmodium* MOI and molecular epidemiological patterns by long amplicon ultra-deep sequencing. The complexity of infections were similar among age groups, symptoms, genders, transmission settings (spatial heterogeneity), as well as over years (pre- vs. post-scale-up interventions). This study demonstrated that long amplicon deep sequencing is a useful tool to investigate multiplicity and molecular epidemiology of *Plasmodium* parasite infections.

**Keywords:** Multiplicity of infection, Merozoite surface protein 1, Amplicon deep sequencing, Molecular epidemiology, *Plasmodium*, Within-host diversity

*Correspondence: dzhong@uci.edu; guiyuny@uci.edu
[1] Program in Public Health, University of California at Irvine, Irvine, CA 92617, USA
Full list of author information is available at the end of the article

Zhong *et al. Malar J (2018) 17:185*

Page 2 of 14

## Background

Malaria is one of the most common infectious diseases and an important public health problem worldwide. In 2016, there were an estimated 216 million cases and 445,000 deaths of malaria occurred worldwide; and nearly half of the world's population lived in 91 countries and territories are at risk of malaria transmission [1]. The majority of malaria cases and deaths (~90%) occur in sub-Saharan Africa. *Plasmodium falciparum* is the most prevalent malaria parasite in sub-Saharan Africa, while *Plasmodium vivax* is the most widespread human malaria with approximately 2.5 billion people at risk of infection worldwide [2]. *Plasmodium vivax* is a major cause of anaemia in an area where *P. falciparum* and *P. vivax* co-exist [3]. Relapses play an important role in the transmission of *P. vivax* in malaria endemic areas [4].With the scaling up of interventions since 2006, primarily mass distribution of insecticide-treated nets (ITNs), indoor residual spraying (IRS), and artemisinin-based combination therapy (ACT), malaria transmission has declined tremendously in the past decade [5].

The extent of genetic diversity and multiplicity of infection (MOI) is essential in understanding malaria epidemiological patterns, transmission intensity, host immune system, and parasite virulence for the development of anti-malarial vaccine as well as evaluating the impact of malaria control interventions. For example, MOI has been used for inferring disease epidemiology such as detecting parasite clearance rates subsequent to anti-malarial treatment [6] and examining the level of anti-malarial drug resistance [7, 8], the impact of transmission intensity on infection complexity [9], parasite virulence related to anti-malarial vaccine development [10, 11], and in-host ecology of malaria infections [12]. Traditional PCR-based methods, such as microsatellite [13, 14] and merozoite surface protein (*msp*) genotyping [14–17], for assessing MOI estimation can lack both sensitivity and specificity, resulting in the apparent problem of underestimating disease complexity [18–20]. Compared to genotyping methods, amplicon deep sequencing provides a rapid, robust, high-throughput approach to detect sequence variants and estimate allele frequency by sequencing a genomic region multiple times, sometimes hundreds or even thousands of times [21]. For example, ultra-deep sequencing of amplicons from the ribosomal, mitochondrion, and apicoplast encoded genes revealed a large complexity of coinfections with an unexpectedly high MOI in *Plasmodium ovale* and *Plasmodium malariae* infections in the endemic areas of Gabon [22]. Use of length polymorphic genes such as *msp2* in amplicon deep sequencing has been shown to display greater sensitivity in detecting minority clones [23].

Using *pvmsp1* short amplicon deep sequencing, Lin et al. [15] identified 67 unique haplotypes from 78 Cambodian *P. vivax* samples with an average MOI of 3.6 within each individual. Over half of the recurrent infections were detected as relapse. Compared to the standard PCR based method, next-generation sequencing revealed up to sixfold higher MOI in *Plasmodium* infections [12, 21]. This technology has unquestionably advanced our understanding of the genetics and evolution of multiclonal infection. However, in previous studies, most of amplicon deep sequencing was performed on two platforms, 454/Roche or Ion Torrent with high error rate and short reads due to technological limitation. By contrast, the Illumina MiSeq/HiSeq platform can generate reads of up to 600 bp length with lower sequencing error rate.

*Plasmodium* merozoite surface protein 1 (*msp1*) is a highly abundant and the most polymorphic antigen, which has been extensively studied in the parasite population [24–26]. *Plasmodium falciparum* has seven variable blocks that are separated either by conserved or semi-conserved regions. The variable block 2 of *pfmsp1* is the most polymorphic region of the antigen [27]. *Plasmodium vivax* has nine variable regions that are separated by 10 interspecies conserved or intraspecies conserved blocks [28]. The variable block 18, located in 42 kDa region of *pvmsp1*, has been identified to be the most polymorphic part of the antigen [11]. These polymorphic regions could be good candidate markers for multiclonal detection of *Plasmodium* infection.

The present study was designed to address the following questions: (1) how useful is amplicon ultra-deep sequencing for determining multiplicity of *Plasmodium* infection and identifying *P. vivax* relapse? (2) is there any difference in multiplicity of *Plasmodium* infection between patients of different symptoms, ages, genders, time, and transmission settings? (3) does intensified intervention since 2006 affect MOI? To address the first question, different lengths of *P. vivax* amplicons and microsatellites for MOI and relapse estimation were compared. For the second and third questions, different groups of *P. falciparum* and *P. vivax* infected patients were compared.

## Methods

### Study site and sample collection

*Plasmodium vivax* parasite samples were collected from 2011 to 2015 in two localities in Ethiopia, including Jimma (351 km away from Addis Ababa) and Asendabo (50 km away east of Jimma) located in Jimma zone of Oromia Regional state in southwestern Ethiopia. The study sites have high elevations, ranging from 1680 to 2010 m above sea level. Malaria transmission is seasonal and unstable, causing frequent epidemics in these areas

Zhong *et al. Malar J* (2018) 17:185

Page 3 of 14

[29–32]. Clinical samples of *P. vivax* from Jimma health centres or hospitals were collected during the peak transmission season (September–November) of 2014-2015, while *P. vivax* samples from Asendabo were collected from Arenado health centre and asymptomatic school children from 2011 to 2012. A total of 138 *P. vivax* samples were used in the study (Table 1). Among them, five patients from Jimma were detected with recurrent infection within 44–73 days. Both the initial and recurrent infected samples were collected from these five patients and included in this study. *Plasmodium falciparum* parasite samples were collected from symptomatic hospital patients or asymptomatic schoolchildren from 2005 to 2015 in lowland and highland areas of western Kenya (Table 1). The study sites included Kombewa and Kendu Bay lowland areas, and Iguhu and Kaimosi in highland areas. Malaria transmission in the lowland area is perennial and hyperendemic, while transmission intensity in the highland is mesoendemic [33–37]. All blood samples were obtained by finger-pricking and dried on filter paper as described previously [34, 38–41].

## PCR amplification and deep sequencing of *pvmsp1* and *pfmsp1*

DNA was extracted from dried blood spots on filter paper by the Saponin/Chelex method [42] and genomic DNA was eluted in a total volume of 200 µl TE buffer. Molecular identification of *P. vivax* and *P. falciparum* for each sample was assessed by nested PCR amplification with species-specific primers designed on the small subunit ribosomal ribonucleic acid (18S rRNA) genes [43, 44]. PCR products for Illumina Hiseq sequencing were prepared by two-step PCR approach using primers specific for highly variable region of *pvmsp1* [28] and *pfmsp1* [27] genes (Fig. 1). The first PCR used appended 5′ end forward and reverse target gene specific primers, while the second PCR used universal primer with barcode primers appended 5′ end (see Additional file 1). The length of PCR products are 463 bp for *P. vivax* (refer to AF435593), and 395 bp for *P. falciparum* (refer to NC_004330).

PCR amplification of each sample was conducted in a 20 µl reaction mixture containing 2 µl of genomic DNA, 4 µl of 5 × PCR buffer, 1 unit of high fidelity PrimeSTAR® GXL DNA Polymerase (Takara Bio USA, Inc., Mountain View, CA), and 10 pmol of each primer. The laboratory strains *P. vivax* Pakchong (MRA-342G) and *P. falciparum* 3D7 (MRA-102G) were also included as control. Amplification reactions were performed with an initial denaturation at 94 °C for 3 min, followed by 35 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 60 s, with a final 6-min extension at 72 °C. Ten samples from each species were amplified in duplicate with unique barcode for confirmation of amplicon. Amplicons were cleaned and normalized to 1–2 ng/µl concentration using the SequalPrep Normalization Plate Kit (Life Technologies, Carlsbad, California). HiSeq Rapid SBS Kit v2 (with reads up to 2 × 250 bp) was used for library preparation. Multiple samples were pooled and sequenced on the Illumina Hiseq 2500 (384-well plate with dual indexing, UCI Genomics High-Throughput Facility).

## Haplotype determination from deep sequencing

Haplotypes of *pvmsp1* and *pfmsp1* variants were determined by SeekDeep software developed by Bailey lab at University of Massachusetts Amherst (http://baileylab.umassmed.edu/seekdeep). This software uses a

## Table 1 Sample collection of *Plasmodium vivax* in Ethiopia and *Plasmodium falciparum* in Kenya

| Country | Species | Location | Elevation (m) | Year | Endemicity[a] [refer] | Symptoms | n | Reads[b] (> 1 k) |
|---------|---------|----------|---------------|------|----------------------|----------|---|------------------|
| Ethiopia | *P. vivax* | Jimma | 1680–1700 | 2014–2015 | 2.4[PCR] [31] | Sym | 67 | 65 |
| | | Asendabo | 1710–2010 | 2011–2012 | 32.4[MIC] [30] | Sym | 48 | 47 |
| | | Asendabo | 1710–2010 | 2011–2012 | 32.4[MIC] [30] | Asym | 23 | 23 |
| Kenya | *P. falciparum* | Iguhu | 1420–1600 | 2005 | 36.5[MIC] [36] | Sym | 29 | 29 |
| | | Iguhu | 1420–1600 | 2005 | 36.5[MIC] [36] | Asym | 29 | 28 |
| | | Iguhu | 1420–1600 | 2015 | 15.0[MIC] [37] | Asym | 22 | 22 |
| | | Kaimosi | 1600–1700 | 2015 | 11.0[MIC] [35] | Asym | 29 | 27 |
| | | Kendu Bay | 1100–1200 | 2015 | 59.3[PCR] [34] | Asym | 29 | 29 |
| | | Kombewa | 1150–1250 | 2005 | 49.4[MIC] [36] | Sym | 29 | 29 |
| | | Kombewa | 1150–1250 | 2005 | 49.4[MIC] [36] | Asym | 29 | 29 |
| | | Kombewa | 1150–1250 | 2015 | 43.5[MIC] [37] | Asym | 29 | 29 |

*Sym* symptomatic; *Asym* asymptomatic; *n* number of samples

[a] Community asymptomatic parasite prevalence by microscope examination (MIC) or PCR detection (PCR)

[b] Number of samples with total joined sequence reads more than 1000

Zhong *et al. Malar J* (2018) 17:185

Page 4 of 14
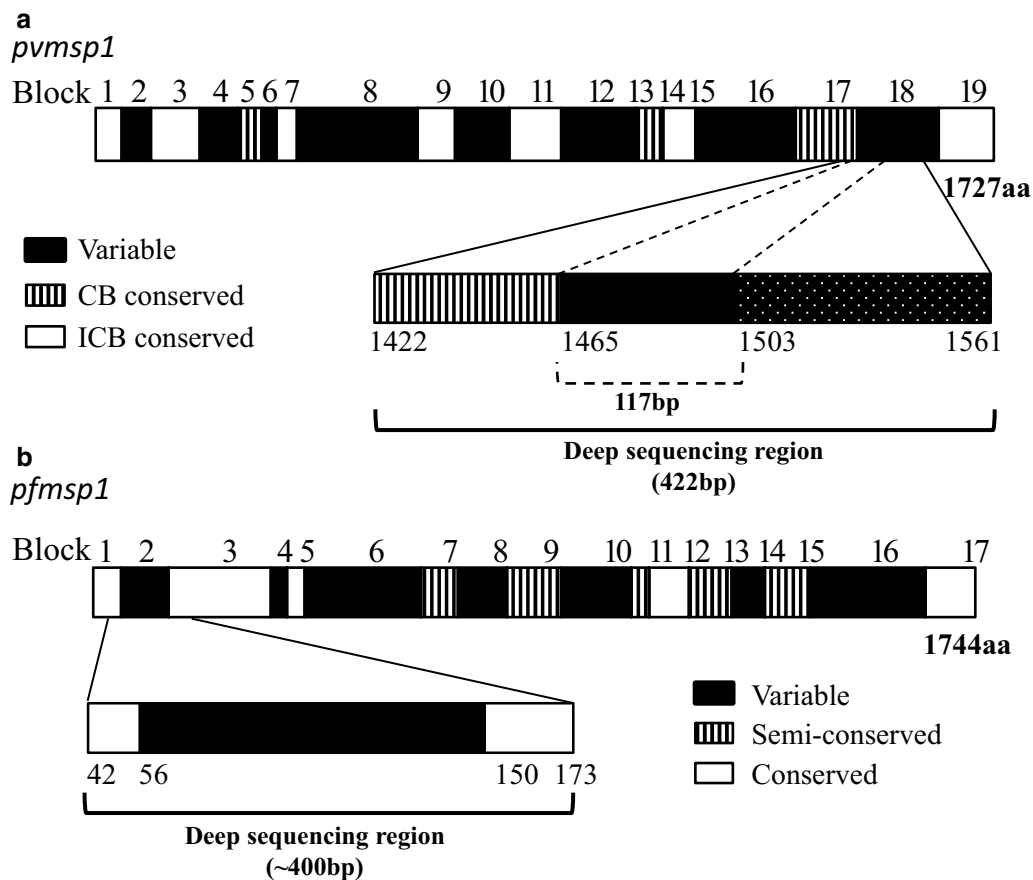


**Fig. 1** Schematic diagrams of the *msp1* protein and amplicon deep sequencing region. **a** *pvmsp1* blocks represent interspecies conserved blocks (ICBs)(white), conserved blocks (CBs) (hatched) and variable regions (black) according to del Portillo et al. [28]; **b** *pfmsp1* blocks represent conserved regions (white), semi-conserved regions (hatched) and variable regions (black) according to Tanabe et al. [27]

clustering method to construct the most likely haplotypes within a patient while removing false haplotypes due to PCR or sequencing error [15]. Before running data on SeekDeep software, all paired-end reads were merged using Fastq-join software with the parameters: Number of percent maximum difference = 8, Number of minimum overlap = 30. Joined reads of each sample were grouped into different clusters after trimming of barcodes, tags, and primers. For each sample, haplotype clusters were determined by within-host reads cutoff frequency at 2.0%. EstimateS (v 9.1.0) program [45] was used to infer estimates of allelic richness. Sample-based rarefaction (haplotype accumulation) curves were plotted with 95% confidence intervals. The input matrix used *msp1* haplotype abundance or incidence data for a set of related samples. Relapse or reinfection of *P. vivax* was classified based on previously published method [15].

## Comparison of amplicon deep sequences of different length and microsatellite genotyping

To compare longer and short sequencing fragment in MOI determination, 117 bp fragment of *pvmsp1* was extracted, which had the same size as the ones previously used for amplicon deep sequencing [15]. MOI was evaluated using the same procedure as describe above. Our previous microsatellite genotyping data of *P. vivax* [39] were also included for comparison in the study. For *P. falciparum*, the shorter amplicon was not able to be extracted for comparison, due to large length difference and extensively polymorphism in the amplicon region. Analysis of variance (ANOVA) and mean comparisons were performed using the JMP statistical software

Zhong *et al. Malar J* (2018) 17:185

Page 5 of 14

package (JMP 12.2.0; SAS Institute, Cary, NC). Mean MOI was compared using the Tukey–Kramer HSD test (alpha = 0.05) or Student's t test.

## Sequence variation analysis and haplotype relationship within multiple infections

MAFFT v7 online version (https://mafft.cbrc.jp/alignment/server/) was used to align DNA sequences [46]. Bioedit v7 was used to calculate sequences identity [47]. Analysis of haplotype and nucleotide diversity was performed by using DnaSP v5 [48]. The Nei's unbiased expected heterozygosity (He) was calculated as a measure of overall genetic diversity for each genotype method [49]. Analysis of Molecular Variance (AMOVA) was conducted by GenAlEx 6.5 to estimate sequences variation within- and between infections [50]. The MEGA v7 was used to create a UPGMA phylogenetic tree [51]. The tree was annotated using the online tool iTOL (interactive Tree of Life) v3 program (http://itol.embl.de/index.shtml) [52]. The PopART v1.7 software was used to construct a minimum spanning haplotype network between haplotypes [53].

## Results

### Sequence reads and haplotype determination

A total of 384 PCR reactions (362 samples and 2 controls as well as 20 replicate PCR reactions) were successfully amplified and sequenced, resulting in 166 M total reads, of which 120 M reads (74%) passed filter, including 100 M with Qscores > 30, with an exception of three *P. vivax* samples and two *P. falciparum* samples with less than 1000 reads that were excluded from the analyses. The haplotype clustering threshold was determined by a subset of samples with replicate PCR reactions. These samples were analysed separately by SeekDeep and the results were compared among replicates as well as with a single clone lab strain (Pakchong or 3D7) as a positive control sample. A threshold cutoff frequency of 2.0% was determined to provide identical results for minor clonal calling between replicates, instead of 0.5% default threshold cut-off [54].

For the 135 *P. vivax* samples, a total of 11,576,219 joined reads were obtained by the fast-join program, of which 4,657,238 (36.1%) were successfully clustered by SeekDeep with an average of 34,498 reads per sample at within-host cluster frequency > 2.0%. The *pvmsp1* amplicon generated identical 422 bp fragment with 88 haplotypes. Among them, 21 *P. vivax* predominant haplotypes (the clone had the highest frequency within infection) were identified (GenBank acc. MG657437–MG657457, Additional file 2). NCBI nucleotide BLAST search identified that 10 out of the 21 unique haplotypes had a perfect match with sequences from GenBank and

> 99% sequence similarity for the others against distinct sequences from GenBank (see Additional file 3).

For the 222 *P. falciparum* samples, a total of 42,832,457 merged reads were generated, of which 23,187,282 (52.5%) joined reads were clustered with an average of 104,447 reads per sample at within-host cluster frequency > 2.0%. The length of *pfmsp1* amplicons varied from 239 to 410 bp with 307 unique haplotypes. Among them, 99 *P. falciparum* predominant haplotypes were identified (GenBank acc. MG675458-MG675556, Additional file 4). NCBI nucleotide BLAST search identified that 25 out of the 99 unique haplotypes had a perfect match with sequences from GenBank and a range of 84.4–99.7% sequence similarity for the others against distinct sequences from GenBank (see Additional file 5).

### Haplotype diversity and population frequency distribution

The 88 unique *pvmsp1* haplotypes exhibited 73 variable (polymorphic) sites, including 12 singleton variable sites. The average haplotype diversity (Hd) and nucleotide diversity (Pi) were 0.969 and 0.055, respectively. All of the 88 haplotypes could be successfully translated into completed amino acid sequences, resulting into 52 distinct amino acid haplotypes. Eight common nucleotide haplotypes each appeared in at least 10 samples (Fig. 2a, Additional file 2), while nearly half (47/88) of haplotypes appeared in only 1 individual sample with within-host frequency ranging from 3.4 to 100%. Approximately 70% (62/88) of the identified haplotypes were detected only as minority variants (within-host frequency < 20%). Some of these minority variants were detected from multiple samples (H09) and multiple locations (H11 and H14–H16). Other minority variants detected formed part of the mutational path (e.g. H43 and H76) between the more common variants, as depicted in a minimum spanning network based on sequence relatedness, adding support that they are true haplotypes and not a result of PCR or sequencing error (Fig. 3). Polyclonal infections were detected in 62.2% (84/135) of samples, ranging from 2 to 6 clones per sample (Fig. 4a). The Nei's unbiased expected heterozygosity at this locus was He = 0.92, representing an average 92% probability for 2 parasite clones harboring different *pvmsp1* haplotypes in the population. Estimates of allelic richness in all the 135 *P. vivax* samples indicated that there was no clear plateau in accumulation curves (Fig. 5a), suggesting more haplotypes would be expected to occur from increased samples.

For the 307 unique *pfmsp1* haplotypes, a total of 20 amplicons with various fragment length were identified, of which the three amplicon sizes 266, 311, and 338 bp appeared in at least 10% of haplotypes (see Additional file 4). All 307 haplotypes could be successfully translated into completed amino acid sequences,

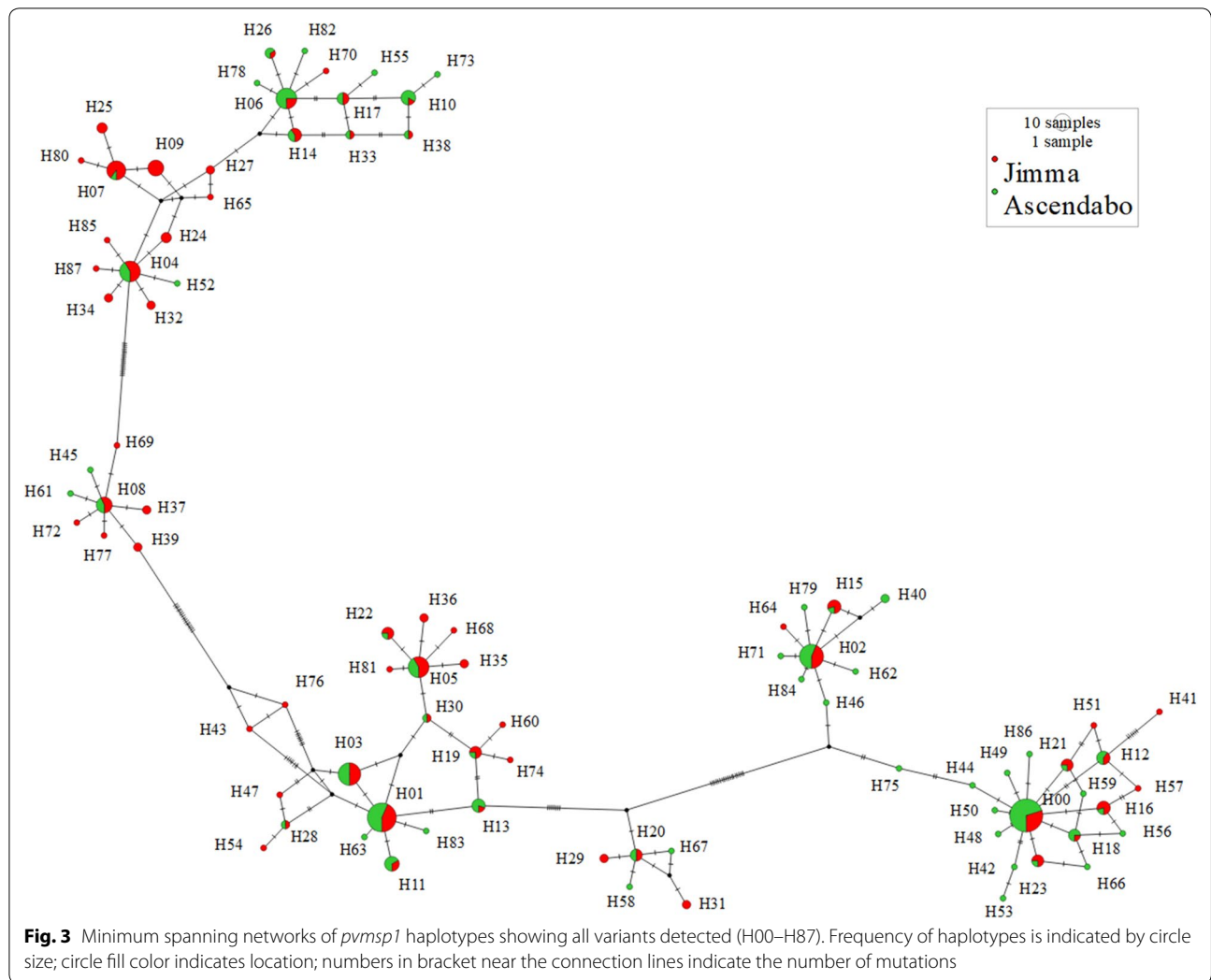Zhong *et al. Malar J* (2018) 17:185

Page 6 of 14
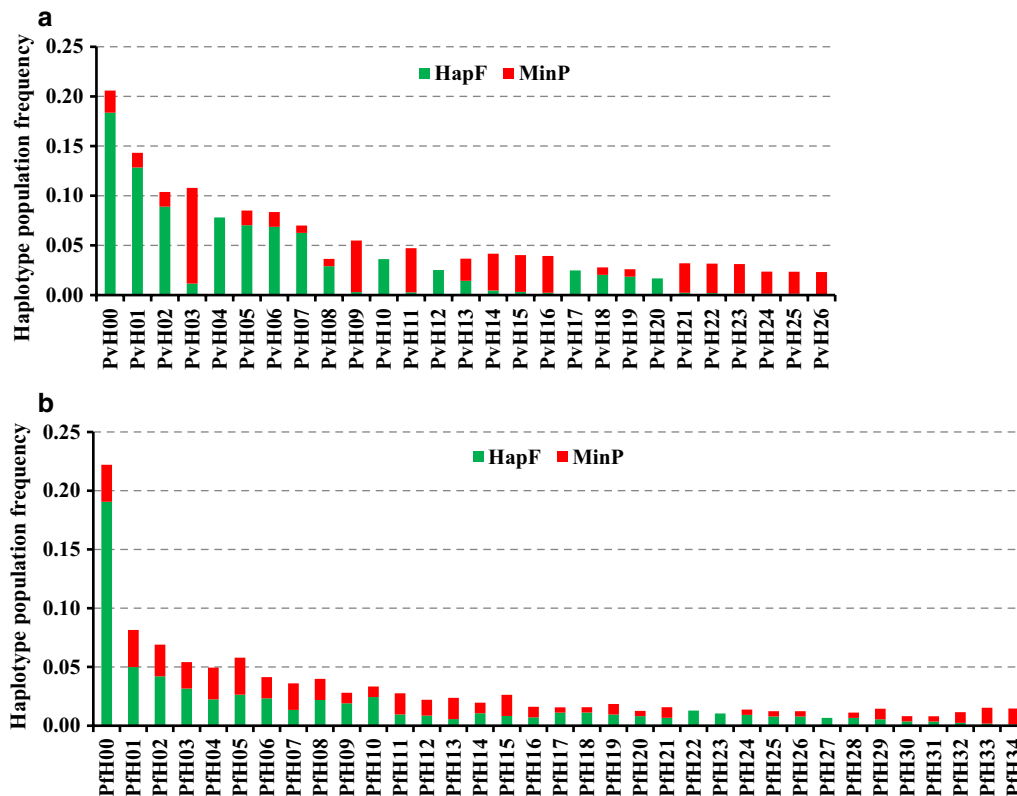


**Fig. 2** Frequency of unique *msp1* haplotypes within the study population. Only haplotypes that appeared in more than 2 samples are shown. The green bar (HapF) represent the frequency of haplotype and the red portions of the columns (MinP) represent the proportion that occurred as a minority variant (existing at 0.5–20% frequency within the individual sample). **a** *P. vivax* population; **b** *P. falciparum* population

resulting in 262 distinct amino acid haplotypes. Eight common nucleotide haplotypes each appeared in at least 10 samples (Fig. 2b, Additional file 4), while nearly 80% (243/307) of haplotypes appeared in only 1 individual sample with within-host frequency ranging from 3.4 to 100%. Approximately half (161/307) of the identified haplotypes were detected only as minority variants. Some of these minority variants were detected from multiple samples and multiple locations. NCBI nucleotide BLAST search identified that 9 out of the 161 minority haplotypes had a perfect match with sequences from GenBank and a range of 91.5–99.7% sequence similarity for the others against distinct sequences from GenBank (Additional file 5). Polyclonal infections were detected in 74.8% (166/222) samples, ranged from 2 to 9 clones per sample (Fig. 4b). The Nei's unbiased expected heterozygosity at this locus was He = 0.95, indicating an average 95% probability to get 2 parasites clones with different *pvmsp1* haplotypes from the population. Similar to *P. vivax* samples, there was no clear asymptote

in accumulation curves from the estimates of allelic richness in all the 222 *P. falciparum* samples (Fig. 5b), suggesting more haplotypes might be found from increased samples.

## Comparison of three methods for MOI and relapse determination in *P. vivax* infection

To examine whether longer amplicon fragment is better for MOI detection and relapse identification in *P. vivax*, the amplicon length was extended from 117 bp [15] to 422 bp for deep sequencing. The results indicated that a longer amplicon had significantly higher MOI (2.16) than that of short amplicon deep sequencing (1.64), and microsatellite markers (1.07) ($F_{2,324} = 27.1$, $P < 0.0001$) (Table 2). Multiclonal infections were detected in 62.2% samples by long fragment deep sequencing, while 45.9% by short fragment deep sequencing and only 5.2% by microsatellite markers. Likewise, number of haplotypes (allele) is higher with a long fragment than that of a short fragment and microsatellites (Table 2). Among the five patients who were detected with recurrent *P. vivax*

Zhong *et al. Malar J* (2018) 17:185

Page 7 of 14



**Fig. 3** Minimum spanning networks of *pvmsp1* haplotypes showing all variants detected (H00–H87). Frequency of haplotypes is indicated by circle size; circle fill color indicates location; numbers in bracket near the connection lines indicate the number of mutations

infection, four of them were identified as relapse infection and one as indeterminate by long amplicon deep sequencing (Fig. 6a). Phylogenetic analysis of haplotypes in the relapse patients indicated that haplotypes were identical between initial and relapse infections, except patient 3 (Pat3) who had an extra minor clone in the initial infections (Pat3-in-H013-8.3) (Fig. 6b). Patient 1 had a different haplotype between initial (Pat1-in-H002-100) and recurrent infections (Pat1-re-H029-100), indicating a new infection or indeterminate.
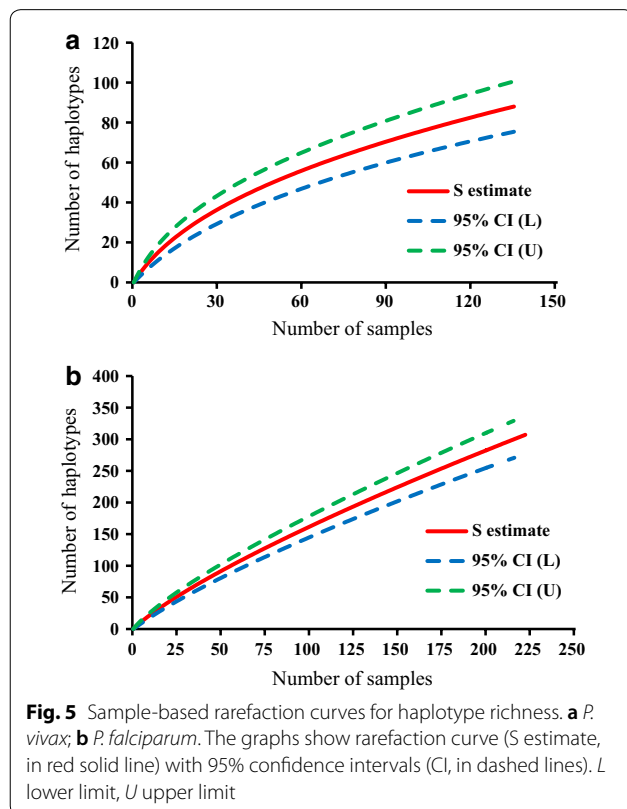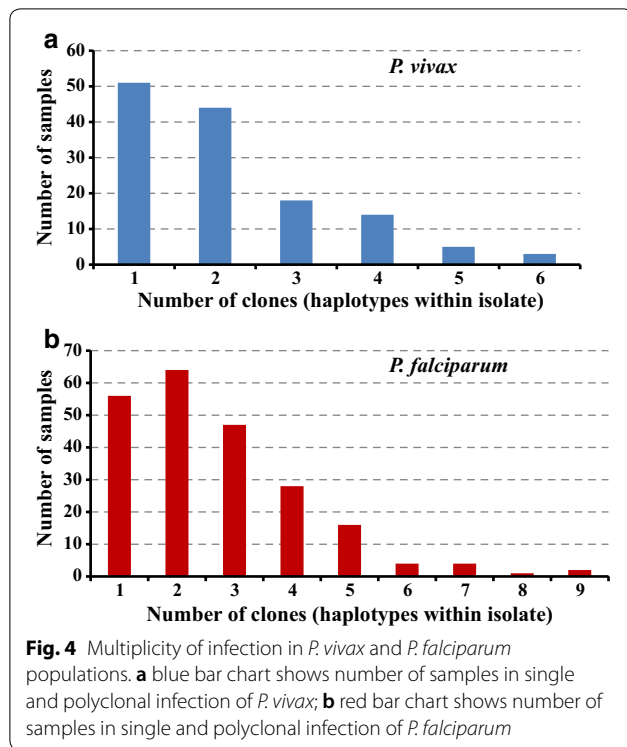
**Multiplicity of *P. vivax* infections in different patient groups**

For the *P. vivax* samples collected from Asendabo in 2011–2012, the percentage of polyclonal infection was 52.2% for asymptomatic and 55.3% for symptomatic patients, respectively. The average MOI in symptomatic patients (n=47) was 2.06±018, which was slightly higher than that of the asymptomatic individuals (MOI=1.83±0.26, n=23), though the difference

was not significant (ANOVA, $F_{1,68}=0.570$, $P=0.453$) (Fig. 7a). Among the symptomatic *P. vivax* samples collected from Jimma in 2014–2015, there was no significant difference in MOI between children and adults (ANOVA, $F_{1,63}=0.212$, $P=0.647$) as well as between females and males (ANOVA, $F_{1,63}=0.185$, $P=0.668$) (Fig. 7b and 7c). The percentage of polyclonal infection ranged from 65.7 to 76.7%, which was slightly higher than that of patients from Asendabo.

**Multiplicity of *P. falciparum* infections in different patient groups**

No significant difference was observed between asymptomatic and symptomatic *P. falciparum* samples from Kombewa (ANOVA $F_{1,56}=0.032$, $P=0.860$) and Iguhu (ANOVA $F_{1,55}=0.728$, $P=0.397$) collected in 2005 (Fig. 8a). Similarly, no difference of MOI was observed from 2005 to 2015 in both Kombewa, the lowland

Zhong *et al. Malar J* (2018) 17:185

Page 8 of 14



**Fig. 4** Multiplicity of infection in *P. vivax* and *P. falciparum* populations. **a** blue bar chart shows number of samples in single and polyclonal infection of *P. vivax*; **b** red bar chart shows number of samples in single and polyclonal infection of *P. falciparum*



**Fig. 5** Sample-based rarefaction curves for haplotype richness. **a** *P. vivax*; **b** *P. falciparum*. The graphs show rarefaction curve (S estimate, in red solid line) with 95% confidence intervals (CI, in dashed lines). *L* lower limit, *U* upper limit

area (ANOVA, $F_{1,56} = 0.832$, $P = 0.366$) and Iguhu, the western Kenya highland area (ANOVA, $F_{1,48} = 1.647$, $P = 0.206$) (Fig. 8b). To examine MOI difference in different transmission settings, the asymptomatic samples collected from western Kenya highland (Iguhu and Kaimosi) and lowland (Kombewa and Kendu Bay) in 2015 were compared. Overall, no significantly difference of MOI was observed among the four study sites (ANOVA, $F_{3,103} = 1.781$, $P = 0.155$). However, the MOI detected in Kombewa (lowland) was 3.10, which was relatively higher than that of the other three sites (2.54 for Iguhu, 2.37 for Kaimosi, and 2.27 for Kendu Bay) and reached significant difference level at $P < 0.05$ between Kombewa and Kendu Bay when the difference was determined by a Student's t test (Fig. 8c).

## Discussion

Multiplicity of infection (MOI), also termed complexity of infection (COI) is defined as the number of different parasite strains co-infecting a single host. MOI can be a useful indicator of immune status and transmission level. Traditionally, MOI was assessed by PCR genotyping of antigen protein genes (*msp1, msp2,* and *glurp*) and microsatellite markers, which were regarded as the gold standard because of their high polymorphism [22]. However, these methods were unable to distinguish sequence variation among parasite strains and detect minority clones within a host. By using next-generation amplicon deep sequencing, the minority clone could be detected as low as 0.5% within-host infection frequency [6, 15]. In the study, the Illumina HiSeq platform combined with Rapid SBS Kit v2 generated reads up to 500 bp with high coverage depth ($\sim 35$ k$\times$ for *P. vivax* and $\sim 100$ k$\times$ for *P. falciparum*). Compared to a previous study by Lin et al. that employed a 117 bp-fragment of *pvmsp1* short amplicon deep sequencing [15], longer amplicon sequencing, by capturing a greater number of polymorphisms, revealed a higher MOI and improved power to detect multiclonal infections. Interestingly, using microsatellite markers with the same parasite population, multiclonal infections were detected only in 5.2% of the samples with an average MOI of 1.07 [39], significantly lower than that estimated by longer amplicon sequencing (a mean of 2.16 MOI). One possible reason might be the missed genotyping of polyclonal infections in some of the tested samples with microsatellite analysis. Such contrast suggested that transmission intensity may not be low. Together with high relapse as identified in the present study, there could be a much larger *P. vivax* reservoir that sustains continual transmission and makes elimination challenging.

The complexity of infection has been suggested to be associated with ages and symptoms in *Plasmodium*

Zhong *et al. Malar J* (2018) 17:185

Page 9 of 14

**Table 2 Comparison of three methods for determination of multiplicity of infection (MOI) of *P. vivax*: long fragment (422 bp) of *pvmsp1* amplicon deep sequencing, short fragment (117 bp) of *pvmsp1* amplicon deep sequencing, and microsatellite marker genotyping**

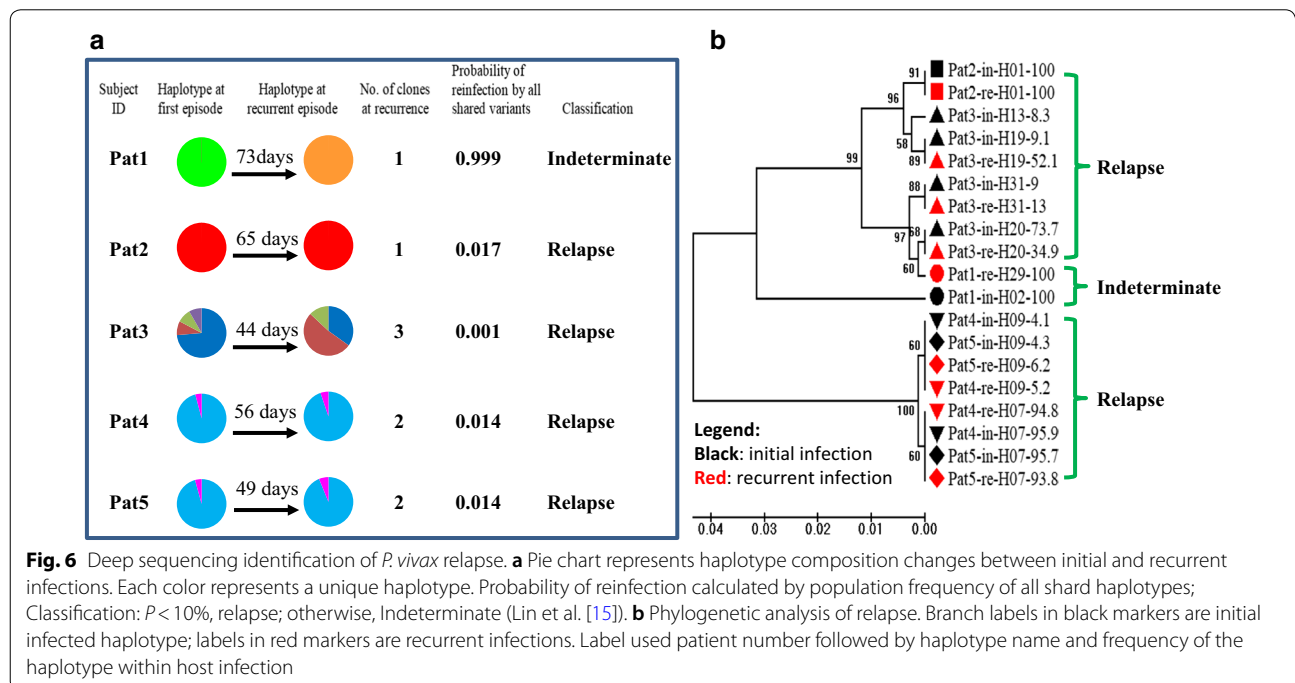|  | Amplicon deep sequencing Long fragment (422 bp) | Amplicon deep sequencing Short fragment (117 bp) | Microsatellite markers[a] |
|---|---|---|---|
| Number of subject | 135 | 135 | 58 |
| Median MOI | 2 | 1 | 1 |
| Mean MOI[a] | 2.16[a] | 1.64[b] | 1.07[c] |
| Max MOI | 6 | 4 | 3 |
| No. polyclonal | 84 | 62 | 3 |
| % polyclonal | 62.2 | 45.9 | 5.2 |
| No. alleles | 88 | 29 | 24 |
| Heterozygosity (He) | 0.92 | 0.84 | 0.77 |

Significant differences were detected in mean MOI among the three methods as indicated by the superscripts (Tukey–Kramer HSD test, $P < 0.05$)
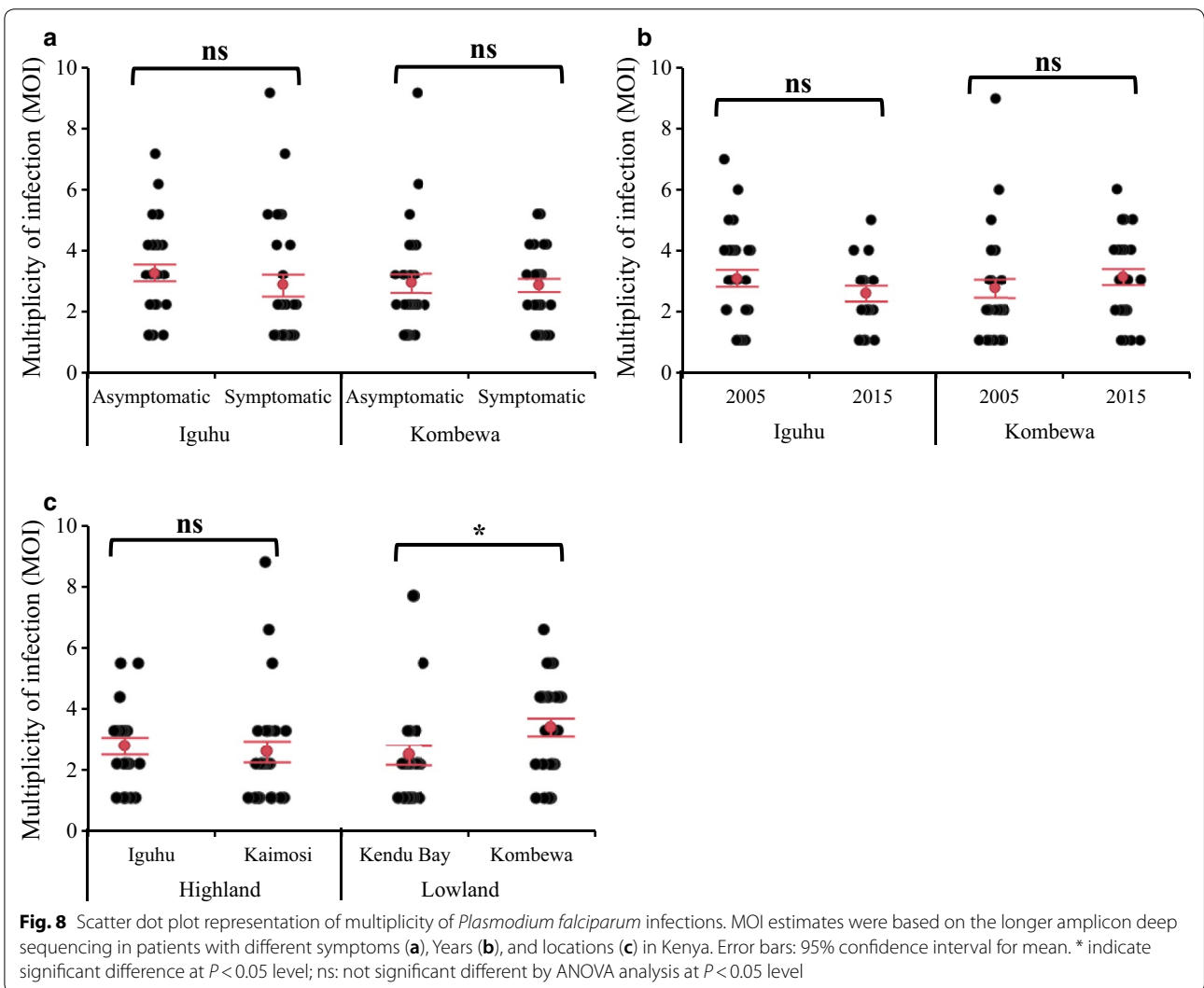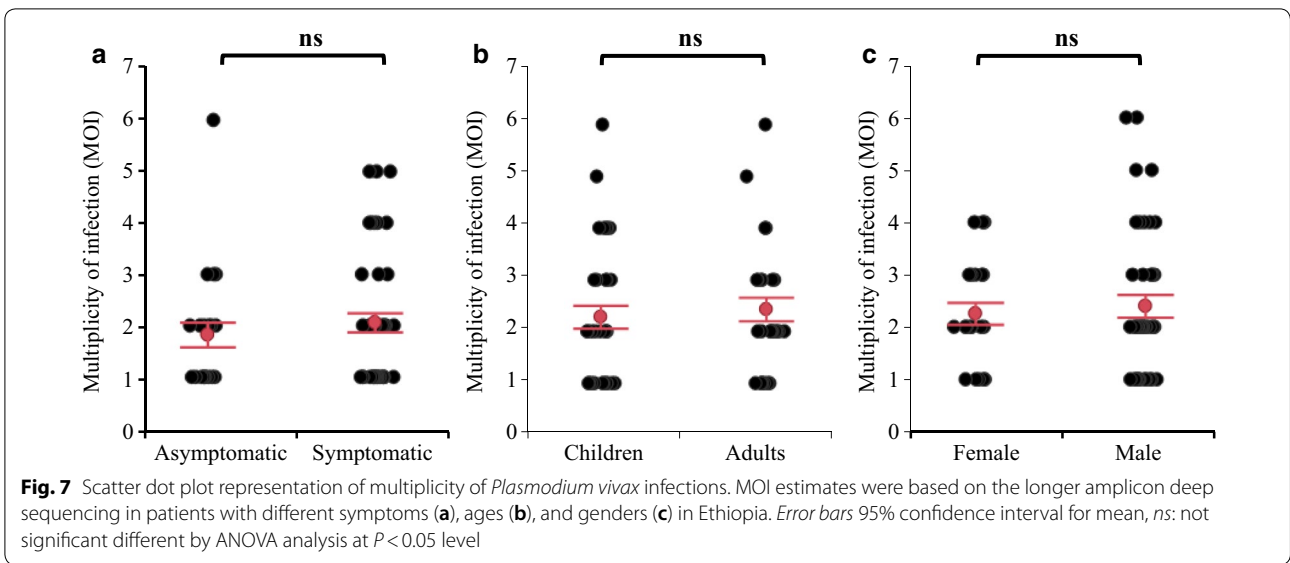
He: Expected heterozygosity corrected for sample size

[a] Refer to Lo et al. (2015) [39]

infections [55–58]. However, in this study, no significant difference was found in *P. vivax* MOI between the symptomatic and asymptomatic infections, adults and children, as well as between male and female groups. Similar patterns were also reported in other studies [59, 60]. In

western Kenya, no notable difference was detected in the multiplicity of *P. falciparum* among asymptomatic school children in low transmission areas (highland) and in high transmission areas (lowland) over 10 years. However, in the high transmission areas (lowland), significantly difference in MOI was detected between locations (Kombewa vs Kendu Bay). The temporal changes in complexity of *P. falciparum* infections could be varied by transmission intensity and our findings indicated that multiclonal parasite genotypes could have remained steady over time in high transmission areas. Several researchers have reported correlations between clinical symptoms and higher MOI [60–68] while others did not find any associations [69–71]. Some studies reported that a reduced risk of clinical malaria was associated with multiclonal infections [72–74], while other studies reported that mono-infections and very common genotypes are more likely to develop severe malaria than multiclonal infections [70, 75]. A positive association between the proportion of polyclonal infections and parasite prevalence has been observed in parasite populations from Indonesia [76] and Papua New Guinea [77], while in other studies, no association or negative correlation was found between the rate of polyclonal infections and parasite prevalence [77, 78]. In Ethiopia, reported malaria cases were respectively 2.6 million and 2.2 million in 2011 and 2015, however, proportion of *P. falciparum* increased by 5% from 2011 to 2015 (Zhou unpublished data), indicating a relative weak reduction in transmission. In our study areas in



**Fig. 6** Deep sequencing identification of *P. vivax* relapse. **a** Pie chart represents haplotype composition changes between initial and recurrent infections. Each color represents a unique haplotype. Probability of reinfection calculated by population frequency of all shard haplotypes; Classification: $P < 10\%$, relapse; otherwise, Indeterminate (Lin et al. [15]). **b** Phylogenetic analysis of relapse. Branch labels in black markers are initial infected haplotype; labels in red markers are recurrent infections. Label used patient number followed by haplotype name and frequency of the haplotype within host infection

Zhong *et al. Malar J* (2018) 17:185

Page 10 of 14



**Fig. 7** Scatter dot plot representation of multiplicity of *Plasmodium vivax* infections. MOI estimates were based on the longer amplicon deep sequencing in patients with different symptoms (**a**), ages (**b**), and genders (**c**) in Ethiopia. *Error bars* 95% confidence interval for mean, *ns*: not significant different by ANOVA analysis at $P < 0.05$ level



**Fig. 8** Scatter dot plot representation of multiplicity of *Plasmodium falciparum* infections. MOI estimates were based on the longer amplicon deep sequencing in patients with different symptoms (**a**), Years (**b**), and locations (**c**) in Kenya. Error bars: 95% confidence interval for mean. * indicate significant difference at $P < 0.05$ level; ns: not significant different by ANOVA analysis at $P < 0.05$ level

Zhong *et al. Malar J (2018) 17:185*

Page 11 of 14

**Table 3 Analysis of molecular variance (AMOVA) of *P. vivax* infections using *pvmsp1* deep sequencing**

| Source of variation | df | SS | MS | Est. variance | Variation (%) | P |
|---|---|---|---|---|---|---|
| Among individuals | 83 | 2276.02 | 27.42 | 8.36 | 70.7 | < 0.001 |
| Within individuals | 157 | 543.45 | 3.46 | 3.46 | 29.3 | |
| Total | 240 | 2819.47 | | 11.83 | 100 | |

Kenya, malaria parasite prevalence in school children in the lowland increased from 40 to 45% from 2011 to 2015 while it decreased from 16% in 2011 to 13% in 2015 in the highlands, results also indicated insignificant changes in transmission in the areas [37].

Long amplicon deep sequencing of *msp1* offers a sensitive tool to detect relapse, defines multiclonal-infected samples, and elucidates within-host genetic diversity and parasite relationships among infections [12, 79–82]. In the present study, a close genetic relationship was found among *P. vivax* clones within-hosts, which explained less than ~30% of the total variance when compared to between-host infections (Table 3). This result suggested that *pvmsp1* haplotypes were more genetically similar within than between hosts. Similar pattern was also observed in *P. falciparum* [83]. The close relatedness among the parasite strains within a host could be a result of frequent recombination and/or selection for drug resistant strains. Further investigation is needed to understand the mechanism generating within-host diversity.

In the study, using long amplicon deep sequencing of high polymorphic makers, *pvmsp1* and *pfmsp1*, minority clones were able to be detected in multiclonal infections. However, there are also a few limitations in the study: (1) only a small polymorphic genomic region is amplified, not covered whole genome variants; (2) the threshold for haplotype cluster calls needs to be determined by empirical methods in each study, due to various sequencing error rates in different sequencing platforms and computational strategies; (3) the PCR slippage might be present in early PCR cycle at the microsatellite repeat unit of length polymorphic *pfmsp1* marker, which resulted in increased frequency of minority clones; (4) there was only a subset of samples with replicate PCR. In order to exclude PCR or sequencing errors, it is better to perform experiments in duplicate of all the samples and use appropriate controls in each study to help determine that no false calls are being made; (5) the low percentage of reads was clustered in our clinical samples compared to laboratory strain (> 99%), suggesting DNA template quality is important. This can be improved by removal of host DNA using an enzyme-based DNA degradation method that selectively digests and depletes human DNA contamination from malaria clinical samples [84]. Another limitation is the lack of a mixed infection positive control, especially for *pfmsp1* with the different product fragment lengths.

## Conclusions

Long-amplicon deep sequencing is a powerful, high-throughput, sensitive method in measuring *Plasmodium* MOI and within-host diversity. Multiclonal infections were common among age groups, symptoms, genders, transmission settings, as well as over years (pre- vs. post-scale-up interventions) in *P. vivax* and *P. falciparum* infections. The study demonstrated that long amplicon deep sequencing is a useful tool to investigate multiplicity and molecular epidemiology of *Plasmodium* parasite infections.

## Additional files

**Additional file 1.** List of two-step PCR primers for amplicon deep sequencing in *P. falciparum pfmsp1* and *P. vivax pvmsp1* genes.

**Additional file 2.** Amplicon deep sequencing data of 135 *P. vivax* samples.

**Additional file 3.** Population frequency, sample counts, and GenBank blast results of the 88 haplotypes identified in the 135 *P. vivax* samples.

**Additional file 4.** Amplicon deep sequencing data of 222 *P. falciparum* samples.

**Additional file 5.** Population frequency, sample counts, and GenBank blast result of the 307 haplotypes identified in the 222 *P. falciparum* samples.

**Author details**
[1] Program in Public Health, University of California at Irvine, Irvine, CA 92617, USA. [2] Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA. [3] Department of Medical Laboratory Sciences, Faculty of Health Sciences, Jimma University, Jimma, Ethiopia. [4] Tropical and Infectious Diseases Research Center, Jimma University, Jimma, Ethiopia. [5] Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya. [6] Department of Medical Microbiology, College of Health Sciences, University of Ghana, Accra, Ghana.

Zhong *et al. Malar J* (2018) 17:185

Page 12 of 14

## Publisher's Note

## References

1. WHO. World Malaria Report 2017. Geneva: World Health Organization; 2017.
2. Howes RE, Battle KE, Mendis KN, Smith DL, Cibulskis RE, Baird JK, et al. Global epidemiology of *Plasmodium vivax*. Am J Trop Med Hyg. 2016;95:15–34.
3. Douglas NM, Anstey NM, Buffet PA, Poespoprodjo JR, Yeo TW, White NJ, et al. The anaemia of *Plasmodium vivax* malaria. Malar J. 2012;11:135.
4. White MT, Shirreff G, Karl S, Ghani AC, Mueller I. Variation in relapse frequency and the transmission potential of *Plasmodium vivax* malaria. Proc Biol Sci. 2016;283:20160048.
5. WHO. World Malaria Report. Geneva: World Health Organization; 2016.
6. Mideo N, Bailey JA, Hathaway NJ, Ngasala B, Saunders DL, Lon C, et al. A deep sequencing tool for partitioning clearance rates following antimalarial treatment in polyclonal infections. Evol Med Public Health. 2016;2016:21–36.
7. Ngondi JM, Ishengoma DS, Doctor SM, Thwai KL, Keeler C, Mkude S, et al. Surveillance for sulfadoxine-pyrimethamine resistant malaria parasites in the Lake and Southern Zones, Tanzania, using pooling and next-generation sequencing. Malar J. 2017;16:236.
8. Rao PN, Uplekar S, Kayal S, Mallick PK, Bandyopadhyay N, Kale S, et al. A method for amplicon deep sequencing of drug resistance genes in *Plasmodium falciparum* clinical isolates from India. J Clin Microbiol. 2016;54:1500–11.
9. Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in *Plasmodium falciparum* transmission: precision, accuracy and costs of metrics. Adv Parasitol. 2014;84:151–208.
10. Juliano JJ, Parobek CM, Brazeau NF, Ngasala B, Randrianarivelojosia M, Lon C, et al. Pooled amplicon deep sequencing of candidate *Plasmodium falciparum* transmission-blocking vaccine antigens. Am J Trop Med Hyg. 2016;94:143–6.
11. Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing patterns of selection and geospatial genetic diversity within

two leading *Plasmodium vivax* candidate vaccine antigens. PLoS Negl Trop Dis. 2014;8:e2796.
12. Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, Ariey F, et al. Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. Proc Natl Acad Sci USA. 2010;107:20138–43.
13. de Souza AM, de Araujo FC, Fontes CJ, Carvalho LH, de Brito CF, de Sousa TN. Multiple-clone infections of *Plasmodium vivax:* definition of a panel of markers for molecular epidemiology. Malar J. 2015;14:330.
14. Lin JT, Juliano JJ, Kharabora O, Sem R, Lin FC, Muth S, et al. Individual *Plasmodium vivax msp1* variants within polyclonal *P. vivax* infections display different propensities for relapse. J Clin Microbiol. 2012;50:1449–51.
15. Lin JT, Hathaway NJ, Saunders DL, Lon C, Balasubramanian S, Kharabora O, et al. Using amplicon deep sequencing to detect genetic signatures of *Plasmodium vivax* relapse. J Infect Dis. 2015;212:999–1008.
16. Hussain MM, Sohail M, Kumar R, Branch OH, Adak T, Raziuddin M. Genetic diversity in merozoite surface protein-1 and 2 among *Plasmodium falciparum* isolates from malarious districts of tribal dominant state of Jharkhand, India. Ann Trop Med Parasitol. 2011;105:579–92.
17. Congpuong K, Sukaram R, Prompan Y, Dornae A. Genetic diversity of the *msp-1, msp-2*, and glurp genes of *Plasmodium falciparum* isolates along the Thai–Myanmar borders. Asian Pac J Trop Biomed. 2014;4:598–602.
18. Khaminsou N, Kritpetcharat O, Daduang J, Charerntanyarak L, Kritpetcharat P. Genetic analysis of the merozoite surface protein-1 block 2 allelic types in *Plasmodium falciparum* clinical isolates from Lao PDR. Malar J. 2011;10:371.
19. Messerli C, Hofmann NE, Beck H-P, Felger I. Critical evaluation of molecular monitoring in malaria drug efficacy trials and pitfalls of length-polymorphic markers. Antimicrob Agents Chemother. 2017;61:e01500–16.
20. Bouyou-Akotet MK, M'Bondoukwe NP, Mawili-Mboumba DP. Genetic polymorphism of merozoite surface protein-1 in *Plasmodium falciparum* isolates from patients with mild to severe malaria in Libreville, Gabon. Parasite. 2015;22:12.
21. Koepfli C, Mueller I. Malaria epidemiology at the clone level. Trends Parasitol. 2017;33:974–85.
22. Lalremruata A, Jeyaraj S, Engleitner T, Joanny F, Lang A, Belard S, et al. Species and genotype diversity of *Plasmodium* in malaria patients from Gabon analysed by next generation sequencing. Malar J. 2017;16:398.
23. Lerch A, Koepfli C, Hofmann N, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. BMC Genomics. 2017;18:864.
24. Gilson PR, Nebl T, Vukcevic D, Moritz RL, Sargeant T, Speed TP, et al. Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. Mol Cell Proteomics. 2006;5:1286–99.
25. Lin CS, Uboldi AD, Epp C, Bujard H, Tsuboi T, Czabotar PE, et al. Multiple *Plasmodium falciparum* merozoite surface protein 1 complexes mediate merozoite binding to human erythrocytes. J Bio Chem. 2016;291:7703–15.
26. Baldwin MR, Li X, Hanada T, Liu SC, Chishti AH. Merozoite surface protein 1 recognition of host glycophorin A mediates malaria parasite invasion of red blood cells. Blood. 2015;125:2704–11.
27. Tanabe K, Mackay M, Goman M, Scaife JG. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J Mol Biol. 1987;195:273–87.
28. del Portillo HA, Longacre S, Khouri E, David PH. Primary structure of the merozoite surface antigen 1 of *Plasmodium vivax* reveals sequences conserved between different Plasmodium species. Proc Natl Acad Sci USA. 1991;88:4030–4.
29. Alemu A, Abebe G, Tsegaye W, Golassa L. Climatic variables and malaria transmission dynamics in Jimma town, South West Ethiopia. Parasit Vectors. 2011;4:30.
30. Demissie F, Kebede A, Shimels T, Beyene P. Assessment of public health implication of malaria-geohelminth co-infection with an emphasis on hookworm-malaria anemia among suspected malaria patients in asendabo, southwest Ethiopia. Ethiop Med J. 2009;47:153–8.
31. Zhou G, Yewhalaw D, Lo E, Zhong D, Wang X, Degefa T, et al. Analysis of asymptomatic and clinical malaria in urban and suburban settings of southwestern Ethiopia in the context of sustaining malaria control and approaching elimination. Malar J. 2016;15:250.

Zhong *et al. Malar J* (2018) 17:185

Page 13 of 14

32. Alemu A, Tsegaye W, Golassa L, Abebe G. Urban malaria and associated risk factors in Jimma town, south-west Ethiopia. Malar J. 2011;10:173.

33. Otsyula GM, Andrew G, Guofa Z, Emmanuel M, Noboru M, Guiyun Y. *Plasmodium falciparum* spatial analysis, Western Kenya highlands. Emerg Infect Dis. 2005;11:1571.

34. Lo E, Nguyen K, Nguyen J, Hemming-Schroeder E, Xu J, Etemesi H, et al. *Plasmodium malariae* prevalence and *csp* gene diversity, Kenya, 2014 and 2015. Emerg Infect Dis. 2017;23:601–10.

35. National Malaria Control Programme (NMCP), Kenya National Bureau of Statistics (KNBS). ICF International: Kenya malaria indicator survey 2015. Maryland: NMCP, KNBS, and ICF International; 2016.

36. Zhou G, Afrane YA, Vardo-Zalik AM, Atieli H, Zhong D, Wamae P, et al. Changing patterns of malaria epidemiology between 2002 and 2010 in western Kenya: the fall and rise of malaria. PLoS ONE. 2011;6:e20318.

37. Zhou G, Lee MC, Githeko AK, Atieli HE, Yan G. Insecticide-treated net campaign and malaria transmission in western Kenya: 2003-2015. Front Public Health. 2016;4:153.

38. Lo E, Zhou G, Oo W, Afrane Y, Githeko A, Yan G. Low parasitemia in submicroscopic infections significantly impacts malaria diagnostic sensitivity in the highlands of Western Kenya. PLoS ONE. 2015;10:e0121763.

39. Lo E, Hemming-Schroeder E, Yewhalaw D, Nguyen J, Kebede E, Zemene E, et al. Transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum* in Ethiopia and prevalence of antimalarial resistant genotypes. PLoS Negl Trop Dis. 2017;11:e0005806.

40. Zhong D, Afrane Y, Githeko A, Cui L, Menge DM, Yan G. Molecular epidemiology of drug-resistant malaria in western Kenya highlands. BMC Infect Dis. 2008;8:105.

41. Zhong D, Afrane Y, Githeko A, Yang Z, Cui L, Menge DM, et al. *Plasmodium falciparum* genetic diversity in western Kenya highlands. Am J Trop Med Hyg. 2007;77:1043–50.

42. Bereczky S, Martensson A, Gil JP, Farnert A. Short report: rapid DNA extraction from archive blood spots on filter paper for genotyping of *Plasmodium falciparum*. Am J Trop Med Hyg. 2005;72:249–51.

43. Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, et al. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. Mol Biochem Parasitol. 1993;61:315–20.

44. Singh B, Bobogare A, Cox-Singh J, Snounou G, Abdullah MS, Rahman HA. A genus- and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies. Am J Trop Med Hyg. 1999;60:687–92.

45. Colwell RK, Elsensohn JE. EstimateS turns 20: statistical estimation of species richness and shared species from samples, with non-parametric extrapolation. Ecography. 2014;37:609–13.

46. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

47. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999;41:95–8.

48. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.

49. Anderson TJC, Haubold B, Williams JT, Estrada-Franco§ JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol Biol Evol. 2000;17:1467–82.

50. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics. 2012;28:2537–9.

51. Kumar S, Stecher G, Tamura K. MEGA7: molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33:1870–4.

52. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–5.

53. Leigh JW, Bryant D. popart: full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6:1110–6.

54. Hathaway NJ, Parobek CM, Juliano JJ, Bailey JA. SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. Nucleic Acids Res. 2018;46:e21.

55. Mayor A, Saute F, Aponte JJ, Almeda J, Gomez-Olive FX, Dgedge M, et al. *Plasmodium falciparum* multiple infections in Mozambique, its relation to other malariological indices and to prospective risk of malaria morbidity. Trop Med Int Health. 2003;8:3–11.

56. Engelbrecht F, Tögel E, Beck H-P, Enwezor F, Oettli A, Felger I. Analysis of *Plasmodium falciparum* infections in a village community in Northern Nigeria: determination of *msp2* genotypes and parasite-specific IgG responses. Acta Trop. 2000;74:63–71.

57. Nguyen HV, van den Eede P, van Overmeir C, Thang ND, Hung LX, D'Alessandro U, et al. Marked age-dependent prevalence of symptomatic and patent infections and complexity of distribution of human Plasmodium species in central Vietnam. Am J Trop Med Hyg. 2012;87:989–95.

58. Tukwasibwe S, Mugenyi L, Mbogo GW, Nankoberanyi S, Maiteki-Sebuguzi C, Joloba ML, et al. Differential prevalence of transporter polymorphisms in symptomatic and asymptomatic *falciparum* malaria infections in uganda. J Infect Dis. 2014;210:154–7.

59. Waltmann A, Darcy AW, Harris I, Koepfli C, Lodo J, Vahi V, et al. High rates of asymptomatic, sub-microscopic *Plasmodium vivax* infection and disappearing *Plasmodium falciparum* malaria in an area of low transmission in Solomon Islands. PLoS Negl Trop Dis. 2015;9:e0003758.

60. Pacheco MA, Lopez-Perez M, Vallejo AF, Herrera S, Arevalo-Herrera M, Escalante AA. Multiplicity of infection and disease severity in *Plasmodium vivax*. PLoS Negl Trop Dis. 2016;10:e0004355.

61. Beck S, Mockenhaupt FP, Bienzle U, Eggelte TA, Thompson WN, Stark K. Multiplicity of *Plasmodium falciparum* infection in pregnancy. Am J Trop Med Hyg. 2001;65:631–6.

62. Kiwuwa MS, Ribacke U, Moll K, Byarugaba J, Lundblom K, Farnert A, et al. Genetic diversity of *Plasmodium falciparum* infections in mild and severe malaria of children from Kampala, Uganda. Parasitol Res. 2013;112:1691–700.

63. Kun JF, Schmidt-Ott RJ, Lehman LG, Lell B, Luckner D, Greve B, et al. Merozoite surface antigen 1 and 2 genotypes and rosetting of *Plasmodium falciparum* in severe and mild malaria in Lambarene, Gabon. Trans R Soc Trop Med Hyg. 1998;92:110–4.

64. Nicastri E, Paglia MG, Severini C, Ghirga P, Bevilacqua N. Disease severity, host characteristics and *Plasmodium falciparum* multiple infections in malaria affected travellers returning from Africa. Travel Med Infect Dis. 2007;5:405–6.

65. Nicastri E, Paglia MG, Severini C, Ghirga P, Bevilacqua N, Narciso P. *Plasmodium falciparum* multiple infections, disease severity and host characteristics in malaria affected travellers returning from Africa. Travel Med Infect Dis. 2008;6:205–9.

66. Pava Z, Handayuni I, Trianty L, Utami RAS, Tirta YK, Puspitasari AM, et al. Passively versus actively detected malaria: similar genetic diversity but different complexity of infection. Am J Trop Med Hyg. 2017;97:1788–96.

67. Ranjit MR, Das A, Das BP, Das BN, Dash BP, Chhotray GP. Distribution of *Plasmodium falciparum* genotypes in clinically mild and severe malaria cases in Orissa, India. Trans R Soc Trop Med Hyg. 2005;99:389–95.

68. Rout R, Mohapatra BN, Kar SK, Ranjit M. Genetic complexity and transmissibility of *Plasmodium falciparum* parasites causing severe malaria in central-east coast India. Trop Biomed. 2009;26:165–72.

69. Conway DJ, Greenwood BM, McBride JS. The epidemiology of multiple-clone *Plasmodium falciparum* infections in Gambian patients. Parasitology. 2009;103:1–5.

70. Durand R, Ariey F, Cojean S, Fontanet A, Ranaivo L, Ranarivelo LA, et al. Analysis of circulating populations of *Plasmodium falciparum* in mild and severe malaria in two different epidemiological patterns in Madagascar. Trop Med Int Health. 2008;13:1392–9.

71. Mara SE, Silué KD, Raso G, N'Guetta SP, N'Goran EK, Tanner M, et al. Genetic diversity of *Plasmodium falciparum* among school-aged children from the Man region, western Côte d'Ivoire. Malar J. 2013;2:419.

72. Al-Yaman F, Genton B, Reeder JC, Anders RF, Smith T, Alpers MP. Reduced risk of clinical malaria in children infected with multiple clones of *Plasmodium falciparum* in a highly endemic area: a prospective community study. Trans R Soc Trop Med Hyg. 1997;91:602–5.

73. Bereczky S, Liljander A, Rooth I, Faraja L, Granath F, Montgomery SM, et al. Multiclonal asymptomatic *Plasmodium falciparum* infections predict a reduced risk of malaria disease in a Tanzanian population. Microbes Infect. 2007;9:103–10.

74. Nielsen MA, Staalsoe T, Kurtzhals JA, Goka BQ, Dodoo D, Alifrangis M, et al. *Plasmodium falciparum* variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity. J Immunol. 2002;168:3444–50.

Zhong *et al. Malar J  (2018) 17:185*

Page 14 of 14

75. Amodu OK, Oyedeji SI, Ntoumi F, Orimadegun AE, Gbadegesin RA, Olumese PE, et al. Complexity of the *msp2* locus and the severity of childhood malaria, in south-western Nigeria. Ann Trop Med Parasitol. 2008;102:95–102.

76. Noviyanti R, Coutrier F, Utami RA, Trimarsanto H, Tirta YK, Trianty L, et al. Contrasting transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum*: implications for malaria control and elimination. PLoS Negl Trop Dis. 2015;9:e0003739.

77. Fola AA, Harrison GLA, Hazairin MH, Barnadas C, Hetzel MW, Iga J, et al. Higher complexity of infection and genetic diversity of *Plasmodium vivax* than *Plasmodium falciparum* across all malaria transmission zones of papua new guinea. Am J Trop Med Hyg. 2017;96:630–41.

78. Getachew S, To S, Trimarsanto H, Thriemer K, Clark TG, Petros B, et al. Variation in complexity of infection and transmission stability between neighbouring populations of *Plasmodium vivax* in southern Ethiopia. PLoS ONE. 2015;10:e0140780.

79. Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, Manske M, et al. Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. PLoS ONE. 2012;7:e32891.

80. Bell AS, de Roode JC, Sim D, Read AF. Within-host competition in genetically diverse malaria infections: parasite virulence and competitive success. Evolution. 2006;60:1358–71.

81. Bushman M, Morton L, Duah N, Quashie N, Abuaku B, Koram KA, et al. Within-host competition and drug resistance in the human malaria parasite *Plasmodium falciparum*. Proc Biol Sci. 2016;283:20153038.

82. Branch OH, Takala S, Kariuki S, Nahlen BL, Kolczak M, Hawley W, et al. *Plasmodium falciparum* genotypes, low complexity of infection, and resistance to subsequent malaria in participants in the Asembo Bay Cohort Project. Infect Immun. 2001;69:7783–92.

83. Nkhoma SC, Nair S, Cheeseman IH, Rohr-Allegrini C, Singlam S, Nosten F, et al. Close kinship within multiple-genotype malaria parasite infections. Proc Biol Sci. 2012;279:2589–98.

84. Oyola SO, Gu Y, Manske M, Otto TD, O'Brien J, Alcock D, et al. Efficient depletion of host DNA contamination in malaria clinical sequencing. J Clin Microbiol. 2013;51:745–51.