Published in final edited form as:

Infect Genet Evol. 2018 July ; 61: 45–52. doi:10.1016/j.meegid.2018.02.018.

Genetic diversity of the *Plasmodium vivax* phosphatidylinositol 3-kinase gene in two regions of the China-Myanmar border

Huguette Gaelle Ngassa Mbenda^{1, #}, Weilin Zeng^{2, #}, Yao Bai², Faiza Amber Siddiqui¹, Zhaoqing Yang^{2, *}, and Liwang Cui^{1, *}

¹Department of Entomology, Pennsylvania State University, 501 ASI Building, University Park, PA 16802, USA

²Department of Pathogen Biology and Immunology, Kunming Medical University, Kunming, China

Abstract

Artemisinin resistance in *Plasmodium falciparum* was associated with mutations in the propeller domain of the *PfK13* gene and increased phosphatidylinositol-3-phosphate kinase (PfPI3K) activity. Assessment of the genetic diversity of the orthologue of *PfK13* gene *i.e.* *PvK12* in *Plasmodium vivax* field samples from the same hotspots of *P. falciparum* artemisinin resistance revealed a limited genetic diversity of *PvK12*. Following the same logic, we analyzed genetic variations of the *PvPI3K* gene in 188 *P. vivax* field isolates from two geographic locations along the China-Myanmar border. Overall, high genetic diversity of *PvPI3K* was observed; parasites from Yunnan's Tengchong County had higher genetic diversity than those from Laiza Township, Kachin State, Myanmar. Almost all the neutrality tests applied detected statistically significant deviation from zero. The negative Tajima's D values in both populations implicated that *PvPI3K* gene might have experienced either a directional selection or an expansion in the population size. There was low linkage disequilibrium between the *PvPI3K* mutations in both populations, suggesting the existence of large, almost panmictic, parasite populations that enabled effective recombination. This later result was confirmed by the detection of a minimum of five recombination events in each population with two major breakpoints. Multiple tests for selection confirmed a signature of purifying selection on *PvPI3K*. All the amino acid mutations were predicted to be neutral for the PI3K protein's function. These findings provide insights on the genetic diversity of *P. vivax* populations along the China-Myanmar border.

Keywords

Artemisinin resistance; *Plasmodium vivax*-PI3K; Southeast Asia; Genetic diversity; China-Myanmar border

*Corresponding authors: Zhaoqing Yang (zhaoqingy92@hotmail.com), Liwang Cui (luc2@psu.edu), Tel.: +1 8148637663, Fax: +1 8148653048.

#These authors contributed equally to the work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Malaria remains a serious public health problem in the Greater Mekong Subregion (GMS) of Southeast Asia (Delacollette et al., 2009; Feng et al., 2015). Malaria transmission in this region is concentrated along international borders (Xu et al., 2016). One of the great challenges for malaria elimination in the GMS lies in the recent emergence of artemisinin resistance in *Plasmodium falciparum* (Noedl et al., 2008; Phyo et al., 2012; Woodrow and White, 2017). In addition, the increasing proportion of *Plasmodium vivax*, which has a latent hypnozoite stage that causes relapses of the disease, also poses challenges to malaria elimination in this region (Howes et al., 2016; Sattabongkot et al., 2004). Since effective chemotherapy is crucial for malaria control, the appearance of drug resistance in *P. vivax* in some countries (Chehuan et al., 2013; Marques et al., 2014; Price et al., 2014) makes it essential to monitor drug resistance in both parasite species. With the indication of reduced CQ efficacy for treating *P. vivax* malaria in the GMS (Yuan et al., 2015), reinforced vigilance on resistance towards other drugs in *P. vivax* is therefore compulsory as advocated (Popovici and Menard, 2015).

Understanding the genetic diversity of genes associated with antimalarial drug resistance in low transmission foci such as Southeast Asia, where resistance to most antimalarial medicines has originated in *P. falciparum*, is necessary for resistance management. In contrast to *P. falciparum* where markers for molecular surveillance of drug resistance are available, no reliable molecular markers associated with *P. vivax* drug resistance have been identified yet. The recent incrimination of mutations in the propeller domain of the *Pfk13* gene as the determinant of artemisinin resistance in *P. falciparum* (Ariey et al., 2014; Ashley et al., 2014; Miotto et al., 2015) has led to the investigations of polymorphisms in *PvK12* gene, orthologue of *Pfk13* in *P. vivax* (Deng et al., 2016; Popovici et al., 2015; Wang et al., 2016). In *P. falciparum*, artemisinin-conferring *Pfk13* mutations are associated with increased PI3K activity and higher levels of the product phosphatidylinositol-3-phosphate (PI3P) (Mbengue et al., 2015). However, no polymorphism within the *PfPI3K* gene in *P. falciparum* field isolates has been correlated with the *Pfk13* polymorphisms that mediate artemisinin resistance. Given the potential involvement of PI3K in drug resistance in *Plasmodium*, we studied the genetic diversity of *PvPI3K* (orthologue of *PfPI3K*) in 188 *P. vivax* parasites sampled from two regions along the China-Myanmar border, where *P. vivax* transmission persists.

2. Materials and methods

2.1. Samples collection

Clinical *P. vivax* samples were collected from 188 patients with acute *P. vivax* malaria attending malaria clinics located along the China-Myanmar border: 96 in Laiza Township, Kachin State, Myanmar in 2015 and 92 in Tengchong County, Yunnan Province, China in 2012-2013 (Fig. 1). Finger-prick blood samples of microscopy-confirmed *P. vivax* cases were spotted onto Whatman filter papers. Informed consent was obtained from patients or their guardians. Ethical approvals were obtained from the institutional review boards of Kunming Medical University and Pennsylvania State University.

2.2. DNA extraction, PCR-amplification and sequencing

Genomic *Plasmodium* DNA was extracted from filter papers using QIAamp DNA Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Genotyping of two polymorphic genes (*msp3α* and *msp3β*) in the parasite samples by PCR/RFLP was done not only to reconfirm the *vivax* species, but also to distinguish single from mixed-strain infections (Bruce et al., 1999; Yang et al., 2006). The full-length *PvPI3K* gene was amplified in all the monoclonal infected samples by nested PCR using Phanta Max Super-Fidelity DNA Polymerase P505d (Nanjing Vazyme Biotech, China) containing high-fidelity Pfu DNA polymerase with outer primers pi3kP1F (5'-TTCCATCTACCACGTACCAC-3') and pi3kP1R (5'-AAAGCGGTAACCATAACG-3'), and nested primers pi3kP2F (5'-GTAAGCCGTTGCAGCAGAA-3') and pi3kP2R (5'-ATGAAAGAGCCAGCATAGCA-3'). The primary PCR reactions were performed in 25 µl with 1 µl of DNA template, 0.5 µM of each primer, 12.5 µl 2× Phanta Max Buffer, 0.5 µl dNTP mix, 0.5 µl Super-Fidelity DNA Polymerase and 9.5 µl water. For the nested PCR, 1 µl primary PCR product was used as the template. For both primary and nested reactions, the following cycling parameters were used: 5 min at 95°C, 35 cycles at 95°C for 15 sec, 56°C for 15 sec, 72°C for 3.5 min, and final extension for 5 min at 72°C. Sequencing was performed on both strands. Sequences were assembled via the SeqMan program of Lasergene software (DNASTAR, Madison, WI, USA) with manual editing and aligned with the Sal I reference sequence (PVX_080480) using the MegAlign program of Lasergene software (DNASTAR, Madison, WI, USA). Sequences were deposited in the GenBank (accession numbers: ----- to -----).

2.3. Genetic diversity

PvPI3K sequences were aligned with the reference sequence from the Sal I strain of *P. vivax* (PVX_080480) (www.plasmodb.org) using the ClustalW algorithm. All the analyses in this study were done using DnaSP v5 software (Librado and Rozas, 2009) and MEGA7 (Kumar et al., 2016) except specified otherwise. Polymorphisms were estimated by the number of single nucleotide polymorphisms (SNPs). Nucleotide diversity was estimated by calculating π , the average number of pairwise nucleotide differences per site (Tajima, 1983) and θ_w , the number of segregating sites. The haplotype diversity (H_d) that measures the number and frequency of haplotypes in a given population was also estimated.

2.4. Tests of neutrality

In order to evaluate whether the *PvPIK3* gene follows the neutral equilibrium model of molecular evolution, we performed the frequency-based tests: Tajima's D test (Tajima, 1989); Fu and Li's F test (Fu and Li, 1993); Fu and Li's D test (Fu and Li, 1993). The Tajima's D statistic calculates the normalized differences between the two measures of nucleotide diversity θ_w and π (Tajima, 1989). Both the Fu and Li's D and F statistics rely on the differences between the number of polymorphic sites in external branches (polymorphisms unique to an extant sequence) and the number of polymorphic sites in internal phylogenetic branches (polymorphisms shared by extant sequences) (Zhang and Ge, 2007). Negative values of these statistics indicate an excess of low-frequency polymorphisms (signifying directional selection or population size expansion), whereas

positive values indicate an excess of intermediate polymorphisms (signifying balancing selection or population size decline) (Das et al., 2004; Fu, 1997; Fu and Li, 1993; Tajima, 1989). For all the statistical analyses, the P-value of 0.05 was considered as the level of significance. The details of the neutrality tests applied are summarized in the Table 1.

2.5. Inference of selection

To implicate whether natural selection has played any role in the evolution course of the *PvPI3K* gene, the ratio of synonymous mutations per synonymous sites (dS) versus non-synonymous mutations per non-synonymous sites (dN) have been estimated for each population. Under the null hypothesis (neutral model of evolution), dS is expected to be equal to dN. An excess of non-synonymous substitutions (dN>dS or dN/dS ratio greater than 1) can be interpreted as positive selection (Gupta et al., 2015), indicating that replacement substitutions increase fitness of the parasite, whereas a rarity of replacement changes (dN<dS or dN/dS ratio lesser than 1) indicates that purifying selection might be working to remove such substitutions from the gene pool (Poon et al., 2009). Statistical significance of the difference was estimated using the codon based Z-test of selection.

We also used McDonald-Kreitman (MK) test to examine departure from neutrality (McDonald and Kreitman, 1991). The MK test compares the ratio of non-synonymous to synonymous polymorphism within a species (Pn/Ps) and the ratio of non-synonymous to synonymous substitutions between closely related species (dN/dS). *Plasmodium knowlesi* *PI3K* sequence was used as the outgroup in this test. Fisher's exact test was used to assess statistical significance. The null hypothesis of MK test assumes Pn/Ps = dN/dS under neutrality, whereas dN/dS < Pn/Ps signifies negative selection. A sliding window approach was used to identify specific regions of *PvPI3K* that deviate from neutral expectations using a window size of 50 and a step size of 10 bp.

Because selection is frequently directed at a few amino acid sites of a gene and sometimes masked by a purifying selection acting also on the gene (Yang and Nielsen, 2002), we conducted several maximum-likelihood tests in the HyPhy package implemented in the Data Monkey Web Server (<http://www.datamonkey.org>) to determine which specific amino acids were targeted by selection (Poon et al., 2009). Since recombination can break the linkage and could be falsely interpreted as adaptive selection, significant recombination events were tested in the DnaSP program and by genetic algorithm for recombination detection (GARD) (Pond et al., 2006) also implemented in the Data Monkey Web Server before running these tests of selection.

2.6. Genetic association within the *PvPI3K* gene and inter-population genetic differentiation

Pairwise linkage disequilibrium (LD) was used to determine the degree of random association between different mutations within this gene. The correlation coefficient (R^2) between paired alleles was estimated using DnaSP v5.0 and Haploview 4.0, and the significance of each association was determined using the Fisher and χ^2 tests after Bonferroni correction. To estimate the proportion of genetic variance of *PvPI3K* gene due to population subdivision, Wright's fixation index of interpopulation variance in allele

frequencies (F_{st}) was calculated. The haplotypes were constructed from both the populations based on the amino acid substitutions occurring at positions with a minimum allele frequency (MAF) of 4%. In order to find out whether there was any genetic relationship between the *PvPI3K* sequences from the China-Myanmar border area and global *P. vivax* populations, a maximum likelihood phylogenetic tree was computed with 50 bootstraps. Two sister taxa to *P. vivax*, *P. cynomolgi* (PCYB_102640) and *P. knowlesi* (PKNH_1017900) retrieved from PlasmoDB were included as outgroups. The global *P. vivax* populations comprise sequences from six Asian (Thailand, China, Myanmar, Cambodia, India, and North Korea), three American (Peru, Mexico, and Colombia) and two African (Mauritania and Madagascar) countries.

2.7. Prediction of the possible effect of PI3K amino acid substitutions on protein function

To predict if any of the amino acids mutated in field *PvPI3K* sequences could affect the protein structure and function, we mapped these residues on a modeled 3D structure. The homology model of PvPI3K was built based on the structures of the class III PI3K from *Drosophila* (PDB code: 2X6H), human (3HIZ), and mouse (2Y3A) using the multiple threading alignment in I-TASSER (Yang et al., 2015). A confidence score (C-score) for estimating the quality of predicted models by I-TASSER was calculated (Yang et al., 2015). In addition, Web-based software PROVEAN (Protein Variation Effect Analyzer) was utilized to predict the effect of amino acid mutations in PvPI3K. A cutoff equal to -2.5 was used for the prediction of effect on the protein structure and function.

3. Results

3.1. Genetic diversity of PvPI3K gene from the China-Myanmar border isolates

Sequencing of *PvPI3K* genes in 188 *P. vivax* field isolates from the China-Myanmar border revealed a large number of SNPs within each population (28 from Kachin, Myanmar, and 69 from Tengchong, China) when compared to the Sal I reference sequence. In the Kachin samples, 11 of 28 mutations were non-synonymous, whereas in the Tengchong samples, 37 of 69 mutations were non-synonymous (Table 1). Nine mutations (A14V, S190P, K205I, G326D, R375P, L488W, H580N, R581L, and Y673H) have been previously reported in Southeast Asia (*P. vivax* Genome Variation Project in <https://www.malariagen.net>; Table 1), whereas the remaining mutations were newly identified (Table 1). Among these mutations, the S190P mutation was approaching high frequencies (83.3% in Kachin and 84.8% in Tengchong). Despite close proximity of the two sites, the two populations only shared five amino acid mutations (S190P, H256N, S293G, Y673H, and R1277M), whereas 6 and 32 amino acid changes were specific to the Kachin and Tengchong populations, respectively.

We used two population genetic indices for the assessment of nucleotide diversity, *i.e.* the average pairwise nucleotide distance π , and the number of polymorphic sites per site θ_w (Table 2). In both populations, π was lower than θ_w , which reflects that most SNPs were rare alleles. Overall π value in the 188 samples was 0.00073, similar to that from the Tengchong population (0.00076), which was slightly higher than that in the Kachin population (Table 2). At the amino acid sequence level, changes at five positions were the most prevalent (with MAF of at least 4%). Based on these five amino acid mutations, a total

of 10 haplotypes were observed in the 188 parasite isolates (Fig. 1). Significant differences existed in the number of haplotypes and prevalence of individual haplotypes between the two study populations. Tengchong samples had significantly higher haplotype diversity (Table 2), and harbored nine haplotypes out of ten (Fig. 1). Six haplotypes (H1, and H3-7) were shared among the two parasite populations. While H1 was the most prevalent haplotype in the Tengchong samples with 69.6 % prevalence, it was the second most prevalent (30.2%) in the Kachin samples after H6 (33.3%). The Tengchong *P. vivax* population had one unique haplotype (H2) not present in the Kachin population, whereas the Kachin *P. vivax* population had three unique haplotypes (H8-H10) not present in Tengchong.

3.2. Mutations within the homologous 3D model of PvPI3K

The full-length PvPI3K protein consists of three distinct domains: Ca^{2+} -dependent membrane-targeting C2 domain, an accessory PI3Ka domain (probably involved in substrate presentation), and the catalytic PI3Kc domain. To predict the effect of amino acid mutations on protein structure, we mapped the mutated residues on the predicted tertiary structure. The homology model of PvPI3K built using I-TASSER aligned well with PI3K from humans. Most of the mutations were extensively spread in the interdomain regions of the protein (Supplementary Fig. S1). Within the conserved domains, the C2 domain has two mutations (R375P and L431I), both of which are completely buried inside the 3D structure (Supplementary Fig. S1). R375P may be deleterious for the protein structure as the positively charged amino acid arginine is replaced with a nonpolar, aliphatic amino acid proline. The PI3Ka domain has only one mutation (C1055R) that lies just outside a helix, and thus may not be significant from a structural point of view. Interestingly, the predicted PI3K catalytic domain (PI3Kc) is completely conserved. In further support of our structure prediction data, the PROVEAN analysis predicted that all the amino acid mutations detected in this study were non-deleterious (neutral) for the PvPI3K protein structure and function (C-score below or equal to -2.5) (Supplementary Table S1).

3.3. Potential selection on PvPI3K gene

Multiple tests were performed on the near full-length sequences as well as individual blocks of *PvPI3K* to determine whether this gene has been under natural selection. There was a significant departure from zero for almost all the neutrality tests in both populations as well in the pooled population. The only exception was that Tajima's D value was not significant in the Kachin samples. All the neutrality statistic values were negative (Tajima's D, Fu & Li's D^* , and Fu & Li's F^*) in both populations, suggesting the presence of rare alleles at low frequencies in these populations. For the near full-length sequence, dN was significantly lower than dS in all the populations (Table 3), indicating purifying selection on this gene. This was further supported by the negatively selected sites identified by the codon-based tests (Supplementary Table S2). Nine sites (five in Kachin and four in Tengchong) were implicated under purifying selection by different tests (Supplementary Table S2).

MK test was used for comparing intra-species polymorphism (Pn/Ps) and interspecies divergence (K_a/K_s) using sequence from the phylogenetically related species *P. knowlesi*. Even though non-statistically significant, values of $Pn/Ps > K_a/K_s$ were observed in the near full-length genes (data not shown) resulting in a neutrality index > 1 (Table 3). This reveals

accumulation of synonymous substitutions, which could be interpreted as negative selection for maintaining protein function by eliminating all deleterious mutations. However, a sliding window for Ka/Ks ratio obtained by comparing the *P. vivax* sequences to sequence of *P. knowlesi* identified some regions with a Ka/Ks value of > 1 in the *PvPI3K* sequence (Fig. 2), thereby indicating the presence of positive selection within the *PvPI3K* gene. A minimum of five recombination events (Rm) was found within the *PvPI3K* gene in the two parasite populations (Table 1). Two major breakpoints were detected in each population: at the nucleotide positions 822 and 3066 in Kachin; at the nucleotide positions 822 and 3494 in Tengchong (Fig. 3). Since Ka/Ks value of > 1 in the *PvPI3K* sequence was detected in the regions flanked by the major recombination breakpoints, we could not validate the hypothesis of positive selection. The latter observation of very low linkage between SNPs (Fig. 3) reinforced this conclusion.

3.4. Genetic differentiation of populations

Maximum-likelihood phylogenetic analysis yielded a tree concordant with the low genetic differentiation ($F_{st} = 0.075$) between the two parasite populations. This suggests extensive genetic exchanges between the two populations. However, there was a clear distinction of the Kachin-Tengchong populations from other *P. vivax* populations from China and Myanmar (included in the global isolates) (Supplementary Fig. S2).

4. Discussion

Malaria along the international border regions is a major hurdle for malaria elimination in the GMS. Recent surveys showed that malaria is still prevalent in the China-Myanmar border region (Lo et al., 2017; Xu et al., 2016; Zhou et al., 2014). Migration across this border appeared asymmetrical, with parasite being imported from Myanmar to China (Lo et al., 2017). Because human migration plays an important role in the dissemination of malaria infection (Ngassa Mbenda and Das, 2014), active surveillance is thus necessary in these border zones. Population genetic studies of parasite populations in the border regions will facilitate the monitoring of parasite movement and measuring the outcome of malaria interventions (Arnott et al., 2012).

The present study compared genetic diversity of the *PI3K* gene in field *P. vivax* isolates from the two sides of the China-Myanmar border. Our study revealed that *PvPI3K* was more diverse in Tengchong as compared to Kachin. This finding was similar to the one observed with the *PvK12* gene despite limited genetic diversity of this gene (Deng et al., 2016). As Tengchong was supposed to be a “sink” site with heavy importation of the parasites (Lo et al., 2017), this observed higher genetic diversity may indicate multiple origins of the parasite strains. Yet, despite this difference in haplotype diversity between the two populations, they did not show a significant population genetic structure as indicated from the absence of LD. This result suggests that extensive genetic exchanges between these parasite populations might have occurred in this region, an indication of a large parasite population along this international border. This finding highlights that intensified control efforts and strengthened monitoring of cross-border human populations are needed to reduce and eventually eliminate malaria transmission in the border region.

A large number of mutations were detected in *PvPI3K*, reflecting a history of natural selection (Ngassa Mbenda and Das, 2016). Genetic recombination within this gene might have played a role in contributing to the high genetic diversity within populations (Koepfli et al., 2011). However, the results of almost all the neutrality tests were statistically deviating from zero in both the populations, suggesting the occurrence of rare alleles and that the *PvPI3K* gene in this region might have experienced either a positive/purifying selection or an expansion/reduction in population size during its evolutionary course. Inference of natural selection identified that the *PvPI3K* gene was under purifying selection and the absence of mutation in the predicted catalytic domain of this protein suggests that this may be due to functional constraints. Furthermore, prediction of all the *PvPI3K* mutations in our study as neutral without influencing the function of the protein aligns well with this hypothesis. An analogous finding was made for the *PvK12* gene in the same regions (Deng et al., 2016). Similarly, genotypic surveys of *PfPI3K* in association with the *PfK13* C580Y mutation in some regions of GMS are not conclusive although *PfPI3K* and *PI3P* were found to be increased in strains of *P. falciparum* carrying the *PfK13* C580Y mutation (Mbengue et al., 2015). In northwest Thailand for example, no polymorphism was found in *PfPI3K* but a non-synonymous mutation was found in the *PI4K* protein instead, which may influence *PI3P* or other relevant members of the phosphoinositol pathway in a manner conducive to evolutionarily fit resistance (Cerqueira et al., 2017). A further study on Cambodian *P. falciparum* isolates detected mutations in the *PI3K* protein of culture-adapted isolates exhibiting increased ring-stage survival and free of *PfK13*-mutations (Mukherjee et al., 2017). However, due to the lack of statistical power, the authors could not determine whether any of these mutations contributed to artemisinin resistance (Mukherjee et al., 2017). Other kinases might be involved in the mechanism of artemisinin resistance such as the phosphatidylinositol-4-phosphate 5-kinase (*PIP5K*) (Mukherjee et al., 2017) implicated in the synthesis of *PIP2*, a substrate for *PI3K* to produce *PIP3* (Mbengue et al., 2015).

Although artemisinin resistance has evolved in *P. falciparum* in the GMS, there is no indication of artemisinin resistance in *P. vivax*. Therefore, the significance of *PvPI3K* mutations on drug resistance remains to be explored. Nevertheless, this study has identified high levels of genetic diversity of *PvPI3K* gene in the *P. vivax* populations along the international border of China and Myanmar, an area that is pursuing malaria elimination. Whereas the lack of LD between the two populations suggests significant mixing of the parasites occurs, these parasites are genetically distinct from those collected from other places within the GMS. Such a difference may be suitable for tracking parasite origins of introduced parasites.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the talent introduction project of Yunnan province (# 2013HA026) and National Institutes of Health, USA (U19AI089672). Z.Y. was supported by the National Science Foundation of China (U1202226 and 31260508).

References

- Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, Kim S, Duru V, Bouchier C, Ma L, Lim P, Leang R, Duong S, Sreng S, Suon S, Chuor CM, Bout DM, Menard S, Rogers WO, Genton B, Fandeur T, Miotto O, Ringwald P, Le Bras J, Berry A, Barale JC, Fairhurst RM, Benoit-Vical F, Mercereau-Puijalon O, Menard D. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature*. 2014; 505:50–55. [PubMed: 24352242]
- Arnott A, Barry AE, Reeder JC. Understanding the population genetics of *Plasmodium vivax* is essential for malaria control and elimination. *Malaria J*. 2012; 11
- Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Mao S, Sam B, Sopha C, Chuor CM, Nguon C, Sovannaroeth S, Pukrittayakamee S, Jittamala P, Chotivanich K, Chutasmit K, Suchatsoonthorn C, Runchaoen R, Hien TT, Thuy-Nhien NT, Thanh NV, Phu NH, Htut Y, Han KT, Aye KH, Mokuolu OA, Olaosebikan RR, Folaranmi OO, Mayxay M, Khanthavong M, Hongvanthong B, Newton PN, Onyamboko MA, Fanello CI, Tshefu AK, Mishra N, Valecha N, Phyto AP, Nosten F, Yi P, Tripura R, Borrmann S, Bashraheil M, Peshu J, Faiz MA, Ghose A, Hossain MA, Samad R, Rahman MR, Hasan MM, Islam A, Miotto O, Amato R, MacInnis B, Stalker J, Kwiatkowski DP, Bozdech Z, Jeeyapant A, Cheah PY, Sakulthaew T, Chalk J, Intharabut B, Silamut K, Lee SJ, Vihokhern B, Kunasol C, Imwong M, Tarning J, Taylor WJ, Yeung S, Woodrow CJ, Flegg JA, Das D, Smith J, Venkatesan M, Plowe CV, Stepniewska K, Guerin PJ, Dondorp AM, Day NP, White NJ. Tracking Resistance to Artemisinin, C. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med*. 2014; 371:411–423. [PubMed: 25075834]
- Bruce MC, Galinski MR, Barnwell JW, Snounou G, Day KP. Polymorphism at the merozoite surface protein-3alpha locus of *Plasmodium vivax*: global and local diversity. *Am J Trop Med Hyg*. 1999; 61:518–525. [PubMed: 10548283]
- Cerqueira GC, Cheeseman IH, Schaffner SF, Nair S, McDew-White M, Phyto AP, Ashley EA, Melnikov A, Rogov P, Birren BW, Nosten F, Anderson TJC, Neafsey DE. Longitudinal genomic surveillance of *Plasmodium falciparum* malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. *Genome Biol*. 2017; 18
- Chehuan YF, Costa MR, Costa JS, Alecrim MG, Nogueira F, Silveira H, Brasil LW, Melo GC, Monteiro WM, Lacerda MV. In vitro chloroquine resistance for *Plasmodium vivax* isolates from the Western Brazilian Amazon. *Malaria journal*. 2013; 12:226. [PubMed: 23819884]
- Das A, Mohanty S, Stephan W. Inferring the population structure and demography of *Drosophila ananassae* from multilocus data. *Genetics*. 2004; 168:1975–1985. [PubMed: 15611168]
- Delacollette C, D'Souza C, Christophel E, Thimasarn K, Abdur R, Bell D, Dai TC, Gopinath D, Lu S, Mendoza R, Ortega L, Rastogi R, Tantiniemitkul C, Ehrenberg J. Malaria trends and challenges in the Greater Mekong Subregion. *The Southeast Asian journal of tropical medicine and public health*. 2009; 40:674–691. [PubMed: 19842400]
- Deng S, Ruan Y, Bai Y, Hu Y, Deng Z, He Y, Ruan R, Wu Y, Yang Z, Cui L. Genetic diversity of the Pvk12 gene in *Plasmodium vivax* from the China-Myanmar border area. *Malaria journal*. 2016; 15:528. [PubMed: 27809837]
- Feng J, Zhou D, Lin Y, Xiao H, Yan H, Xia Z. Amplification of pfmdr1, pfcr, pvmdr1, and K13 propeller polymorphisms associated with *Plasmodium falciparum* and *Plasmodium vivax* isolates from the China-Myanmar border. *Antimicrobial agents and chemotherapy*. 2015; 59:2554–2559. [PubMed: 25691632]
- Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. 1997; 147:915–925. [PubMed: 9335623]
- Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993; 133:693–709. [PubMed: 8454210]
- Gupta B, Reddy BP, Fan Q, Yan G, Sirichaisinthop J, Sattabongkot J, Escalante AA, Cui L. Molecular Evolution of PvMSP3alpha Block II in *Plasmodium vivax* from Diverse Geographic Origins. *PloS one*. 2015; 10:e0135396. [PubMed: 26266539]
- Howes RE, Battle KE, Mendis KN, Smith DL, Cibulskis RE, Baird JK, Hay SI. Global Epidemiology of *Plasmodium vivax*. *The American journal of tropical medicine and hygiene*. 2016; 95:15–34.

- Koepfli C, Ross A, Kiniboro B, Smith TA, Zimmerman PA, Siba P, Mueller I, Felger I. Multiplicity and Diversity of *Plasmodium vivax* Infections in a Highly Endemic Region in Papua New Guinea. *Plos Neglect Trop D*. 2011; 5
- Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol*. 2016; 33:1870–1874. [PubMed: 27004904]
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25:1451–1452. [PubMed: 19346325]
- Lo E, Lam N, Hemming-Schroeder E, Nguyen J, Zhou G, Lee MC, Yang Z, Cui L, Yan G. Frequent Spread of *Plasmodium vivax* Malaria Maintains High Genetic Diversity at the Myanmar-China Border without Distance and Landscape Barriers. *The Journal of infectious diseases*. 2017
- Marques MM, Costa MR, Santana Filho FS, Vieira JL, Nascimento MT, Brasil LW, Nogueira F, Silveira H, Reyes-Lecca RC, Monteiro WM, Lacerda MV, Alecrim MG. *Plasmodium vivax* chloroquine resistance and anemia in the western Brazilian Amazon. *Antimicrobial agents and chemotherapy*. 2014; 58:342–347. [PubMed: 24165179]
- Mbengue A, Bhattacharjee S, Pandharkar T, Liu H, Estiu G, Stahelin RV, Rizk SS, Njimoh DL, Ryan Y, Chotivanich K, Nguon C, Ghorbal M, Lopez-Rubio JJ, Pfreder M, Emrich S, Mohandas N, Dondorp AM, Wiest O, Haldar K. A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature*. 2015; 520:683–687. [PubMed: 25874676]
- Mcdonald JH, Kreitman M. Adaptive Protein Evolution at the Adh Locus in *Drosophila*. *Nature*. 1991; 351:652–654. [PubMed: 1904993]
- Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, Lim P, Mead D, Oyola SO, Dhorda M, Imwong M, Woodrow C, Manske M, Stalker J, Drury E, Campino S, Amenga-Etego L, Thanh TN, Tran HT, Ringwald P, Bethell D, Nosten F, Phyto AP, Pukrittayakamee S, Chotivanich K, Chuor CM, Nguon C, Suon S, Sreng S, Newton PN, Mayxay M, Khanthavong M, Hongvanthong B, Htut Y, Han KT, Kyaw MP, Faiz MA, Fanello CI, Onyamboko M, Mokuolu OA, Jacob CG, Takala-Harrison S, Plowe CV, Day NP, Dondorp AM, Spencer CC, McVean G, Fairhurst RM, White NJ, Kwiatkowski DP. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nature genetics*. 2015; 47:226–234. [PubMed: 25599401]
- Mukherjee A, Bopp S, Magistrado P, Wong W, Daniels R, Demas A, Schaffner S, Amaratunga C, Lim P, Dhorda M, Miotto O, Woodrow C, Ashley EA, Dondorp AM, White NJ, Wirth D, Fairhurst R, Volkman SK. Artemisinin resistance without pfcKelch13 mutations in *Plasmodium falciparum* isolates from Cambodia. *Malaria J*. 2017; 16
- Ngassa Mbenda HG, Das A. Molecular evidence of *Plasmodium vivax* mono and mixed malaria parasite infections in Duffy-negative native Cameroonians. *PloS one*. 2014; 9:e103262. [PubMed: 25084090]
- Ngassa Mbenda HG, Das A. Analysis of genetic diversity in the chloroquine-resistant gene *Pfcr*t in field *Plasmodium falciparum* isolates from five regions of the southern Cameroon. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2016; 44:450–458.
- Noedl H, Se Y, Schaefer K, Smith BL, Socheat D, Fukuda MM. Artemisinin Resistance in Cambodia 1 Study, C. Evidence of artemisinin-resistant malaria in western Cambodia. *N Engl J Med*. 2008; 359:2619–2620. [PubMed: 19064625]
- Phyo AP, Nkhoma S, Stepniewska K, Ashley EA, Nair S, McGready R, ler Moo C, Al-Saai S, Dondorp AM, Lwin KM, Singhasivanon P, Day NP, White NJ, Anderson TJ, Nosten F. Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *Lancet*. 2012; 379:1960–1966. [PubMed: 22484134]
- Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. GARD: a genetic algorithm for recombination detection. *Bioinformatics*. 2006; 22:3096–3098. [PubMed: 17110367]
- Poon AFY, Frost SDW, Pond SLK. Detecting Signatures of Selection from DNA Sequences Using Datamonkey. *Methods Mol Biol*. 2009; 537:163–183. [PubMed: 19378144]
- Popovici J, Kao S, Eal L, Bin S, Kim S, Menard D. Reduced polymorphism in the Kelch propeller domain in *Plasmodium vivax* isolates from Cambodia. *Antimicrobial agents and chemotherapy*. 2015; 59:730–733. [PubMed: 25385109]

- Popovici J, Menard D. Challenges in Antimalarial Drug Treatment for Vivax Malaria Control. *Trends Mol Med*. 2015; 21:776–788. [PubMed: 26611336]
- Price RN, von Seidlein L, Valecha N, Nosten F, Baird JK, White NJ. Global extent of chloroquine-resistant *Plasmodium vivax*: a systematic review and meta-analysis. *The Lancet Infectious diseases*. 2014; 14:982–991. [PubMed: 25213732]
- Sattabongkot J, Tsuboi T, Zollner GE, Sirichaisinthop J, Cui L. *Plasmodium vivax* transmission: chances for control? *Trends in parasitology*. 2004; 20:192–198. [PubMed: 15099559]
- Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 1983; 105:437–460. [PubMed: 6628982]
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989; 123:585–595. [PubMed: 2513255]
- Wang M, Siddiqui FA, Fan Q, Luo E, Cao Y, Cui L. Limited genetic diversity in the PvK12 Kelch protein in *Plasmodium vivax* isolates from Southeast Asia. *Malaria journal*. 2016; 15:537. [PubMed: 27821166]
- Woodrow CJ, White NJ. The clinical impact of artemisinin resistance in Southeast Asia and the potential for future spread. *FEMS microbiology reviews*. 2017; 41:34–48. [PubMed: 27613271]
- Xu JW, Li Y, Yang HL, Zhang J, Zhang ZX, Yang YM, Zhou HN, Havumaki J, Li HX, Liu H, Zhou H, Xie XY, Dong JX, Zhang Y, Sun XY, Li B, Li JY, Tian YH, Wang PY, Li BF. Malaria control along China-Myanmar Border during 2007-2013: an integrated impact evaluation. *Infectious diseases of poverty*. 2016; 5:75. [PubMed: 27507163]
- Yang ZH, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*. 2002; 19:908–917. [PubMed: 12032247]
- Yang JY, Yan RX, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015; 12:7–8. [PubMed: 25549265]
- Yang Z, Miao J, Huang Y, Li X, Putaporntip C, Jongwutiwes S, Gao Q, Udomsangpetch R, Sattabongkot J, Cui L. Genetic structures of geographically distinct *Plasmodium vivax* populations assessed by PCR/RFLP analysis of the merozoite surface protein 3beta gene. *Acta Trop*. 2006; 100:205–212. [PubMed: 17129568]
- Yuan L, Wang Y, Parker DM, Gupta B, Yang Z, Liu H, Fan Q, Cao Y, Xiao Y, Lee MC, Zhou G, Yan G, Baird JK, Cui L. Therapeutic responses of *Plasmodium vivax* malaria to chloroquine and primaquine treatment in northeastern Myanmar. *Antimicrobial agents and chemotherapy*. 2015; 59:1230–1235. [PubMed: 25512415]
- Zhang LB, Ge S. Multilocus analysis of nucleotide variation and speciation in *Oryza officinalis* and its close relatives. *Molecular Biology and Evolution*. 2007; 24:769–783. [PubMed: 17182895]
- Zhou G, Sun L, Xia R, Duan Y, Xu J, Yang H, Wang Y, Lee MC, Xiang Z, Yan G, Cui L, Yang Z. Clinical malaria along the China-Myanmar border, Yunnan Province, China, January 2011–August 2012. *Emerging infectious diseases*. 2014; 20:675–678. [PubMed: 24656298]

Highlights for review

- High genetic diversity of *PvPI3K* gene along the international China-Myanmar border
- Low linkage disequilibrium between the *PvPI3K* mutations was observed
- Purifying selection was found to play a role in the evolution of the *PvPI3K* gene
- All the amino acid changes were predicted to be neutral for the protein's function
- Tengchong-Kachin *P. vivax* parasites are genetically distinct from other areas' ones

Fig. 1. Distribution of major haplotypes. Major Haplotypes were constructed on the basis of mutations occurring at five amino acid sites with MAF of at least 4%.

Fig. 2. Distribution of Ka/Ks ratio across *PvPI3K* gene. Ka/Ks ratio was calculated across the gene. A sliding window approach (with a window size of 50 and a step size of 10 bp) by comparing the *P. vivax* sequences to the sequence of *P. knowlesi* was used to identify specific regions of *PvPI3K* that deviate from neutral expectations (represented in the figure by the horizontal line).

Fig. 3.

Linkage disequilibrium pattern in *PvPI3K* from Kachin (A) and Tengchong (B). LD analysis was performed after removal of the fragments under recombination events. The strength of statistical significance of LD between a pair of SNPs is represented with the extent of darkness of the boxes (increasing intensity of black color depicts stronger LD, while fading grey and white color show weaker and no LD, respectively).

Table 1
Frequency of nonsynonymous mutations in Kachin and Tengchong

Substitutions	Kachin (n=96)	Tengchong (n=92)	P value*	Total (n=188)
A14V	-	1.09%		0.53%
V41F	-	1.09%		0.53%
R44K	-	1.09%		0.53%
G67D	-	2.17%		1.06%
S176R	-	1.09%		0.53%
S190P	83.33%	84.78%	0.7862	84.04%
G195R	-	1.09%		0.53%
K196M	-	1.09%		0.53%
K205I	-	3.26%		1.60%
R253L	-	1.09%		0.53%
H256N	13.54%	1.09%	0.0013	7.45%
C275Y	-	1.09%		0.53%
S293G	8.33%	13.04%	0.2951	10.64%
D310N	-	1.09%		0.53%
G326D	1.04%	-		0.53%
P329L	-	1.09%		0.53%
S362T	-	2.17%		1.06%
R375P	1.04%	-		0.53%
L431I	-	1.09%		0.53%
G468D	-	2.17%		1.06%
L488W	-	1.09%		0.53%
R555S	3.13%	-		1.60%
V575A	1.04%	-		0.53%
H580N	-	1.09%		0.53%
H580Q	-	4.35%		2.13%
R581L	-	3.26%		1.60%
A588T	-	2.17%		1.06%
R643L	-	1.09%		0.53%
Y673H	28.13%	58.69%	< 0.0001	43.09%
A708S	-	1.09%		0.53%
R721G	-	1.09%		0.53%

Substitutions	Kachin (n=96)	Tengchong (n=92)	P value*	Total (n=188)
R747L	-	1.09%		0.53%
E766K	-	1.09%		0.53%
S786A	-	1.09%		0.53%
C1055R	-	1.09%		0.53%
S1153F	-	1.09%		0.53%
R1157H	3.13%	-		1.60%
E1179K	-	1.09%		0.53%
S1263T	1.04%	-		0.53%
D1266E	-	1.09%		0.53%
R1277M	1.04%	2.17%		1.06%
A1378P	-	1.09%		0.53%
P1504T	-	1.09%		0.53%

The shadowed are the mutations already published in Southeast Asia. The rest are the newly identified in this study

* SNPs with a total minor allele frequency of > 4% were used in comparison between the two sites (χ^2 test)

Table 2
Nucleotide diversity and neutrality tests of *PvPI3K* in *P. vivax* samples from Kachin and Tengchong.

Regions	s	h	Rm	Hd	Θw	π	Tajima's D	Fu & Li's D*	Fu & Li's F*
Kachin (n=96)	28	21	5	0.862	0.00106	0.00066	-1.152	-2.591*	-2.437*
Tengchong (n=92)	69	60	5	0.988	0.00263	0.00076	-2.329**	-5.021**	-4.700**
Total (n=188)	85	76	5	0.958	0.00284	0.00073	-2.293**	-6.412**	-5.434**

s: number of SNPs
h: number of haplotypes
Fu&Li's F*
Hd: haplotype diversity
π: average number of segregating sites
π: average number of pairwise nucleotide differences per site
* : P-value < 0.05
** : P-value<0.02 for Fu&Li's D* and
** : P-value<0.01 for Tajima's D
Rm: minimum recombination events

Table 3

Tests of selection on *PvPI3K* gene from *P. vivax* parasite collected in Kachin and Tengchong.

Populations	MK test		Codon-based Z-test
	NI	P-value	dN-dS
Kachin (n=96)	0.513	0.16 (NS)	-2.072 *
Tengchong (n=92)	1.428	0.18 (NS)	-2.061 *
Total (n=188)	1.224	0.40 (NS)	-2.24 *

NI: Neutrality index

NS: not significant

* : P-value < 0.05

MK: McDonald and Kreitman