Whole-genome sequencing of a *Plasmodium vivax* isolate from the China-Myanmar border area

Hai-Mo Shen[1], Shen-Bo Chen[1], Yue Wang[1,2], Jun-Hu Chen[1]/+

[1]National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Key Laboratory of Parasite and Vector Biology, Chinese Ministry of Health, WHO Collaborating Center for Malaria, Schistosomiasis and Filariasis, Shanghai, People's Republic of China
[2]Institute of Parasitic Diseases, Zhejiang Academy of Medical Sciences, Hangzhou, People's Republic of China

Currently, there is a trend of an increasing number of Plasmodium vivax *malaria cases in China that are imported across its Southeast Asia border, especially in the China-Myanmar border area (CMB). To date, little is known about the genetic diversity of* P. vivax *in this region. In this paper, we report the first genome sequencing of a* P. vivax *isolate (CMB-1) from a vivax malaria patient in CMB. The sequencing data were aligned onto 96.43% of the* P. vivax *Salvador I reference strain (Sal I) genome with 7.84-fold coverage as well as onto 98.32% of 14 Sal I chromosomes. Using the* de novo *assembly approach, we generated 8,541 scaffolds and assembled a total of 27.1 Mb of sequence into CMB-1 scaffolds. Furthermore, we identified all 295 known* vir *genes, which is the largest subtelomeric multigene family in malaria parasites. These results provide an important foundation for further research on* P. vivax *population genetics.*

Key words: *Plasmodium vivax* - genome - *vir*

As the most common human malaria species with the widest geographic distribution, *Plasmodium vivax* is mostly found outside of Africa and is especially prevalent in Southeast Asia and America (Price et al. 2007). The *P. vivax* parasite is now considered the cause of severe malaria syndromes that have been blamed on *P. falciparum* (Price et al. 2009). It was estimated that half of the world's population is at risk of *P. vivax* malaria (Guerra et al. 2010).

In China, *P. vivax* was the major species for a relatively long time. The Yunnan province remains the highest transmission area in China, particularly in the southern border areas adjacent to Myanmar, a highly endemic area for *P. vivax* malaria in the Greater Mekong Subregion countries (Zhou et al. 2014). Due to the increasing numbers of Chinese working abroad, the number of imported *P. vivax* cases has exhibited an increasing trend in recent years (Feng et al. 2015). The imported *P. vivax* malaria may lead to high malaria risk in malaria-free localities where the *Anopheles sinensis* mosquito is prevalent, particularly in central China, such as in Anhui and Henan provinces (Gao et al. 2004). In 2012, 1,143 *P. vivax* malaria cases were reported in China, accounting for 41.9% of the total malaria cases.

Hundreds of *P. falciparum* isolates have been sequenced or genotyped (Winzeler 2008), but less infor-

mation on *P. vivax* isolates has been reported. The first complete genome of *P. vivax* was published in 2008 (Carlton et al. 2008), revealing that *P. vivax* resembles other *Plasmodium* spp in gene content and metabolic traits, but it possesses novel gene families and potential alternative invasion pathways not previously been recognised. Although they were started more recently, there are several *P. vivax* isolate genome sequencing projects underway and more sequence data have been made available (Neafsey et al. 2012). Previously, 23 isolates genome data were accessible in the National Center for Biotechnology Information (NCBI) database. However, no isolates from China or China's borders with Southeast Asian countries have previously been included.

In this paper, we report the first *P. vivax* genome sequence of a clinical isolate obtained in the China-Myanmar border area (CMB-1) as well as in the China-Southeast Asia border areas. The genomic DNA of the *P. vivax* CMB-1 for sequencing was extracted from a whole blood sample from a patient with microscopically positive for *P. vivax* and polymerase chain reaction confirmed sole infection with *P. vivax*. The Ethical Committee of the National Institute of Parasitic Diseases, China Center for Disease Control and Prevention approved the study (NIPD 2013-010). The study protocol, potential risks and potential benefits were explained to the patient and informed consent was verbally obtained. The genomic DNA was used to construct the Illumina sequencing library with insert sizes of 360 bp. The library was sequenced on a HiSeq 2000 sequencer. After filtering the sequences as for the *Homo sapiens* genome, the reads were *de novo* assembled using an A5 assembly pipeline (Coil et al. 2015). The Illumina sequencing reads have been submitted to the NCBI Short Read Archive (accession SRS941624).

The whole genome sequencing generated 31,471,932 paired-end reads with an average read length of 125 bp. Low-quality bases and adapters were trimmed using Trim-
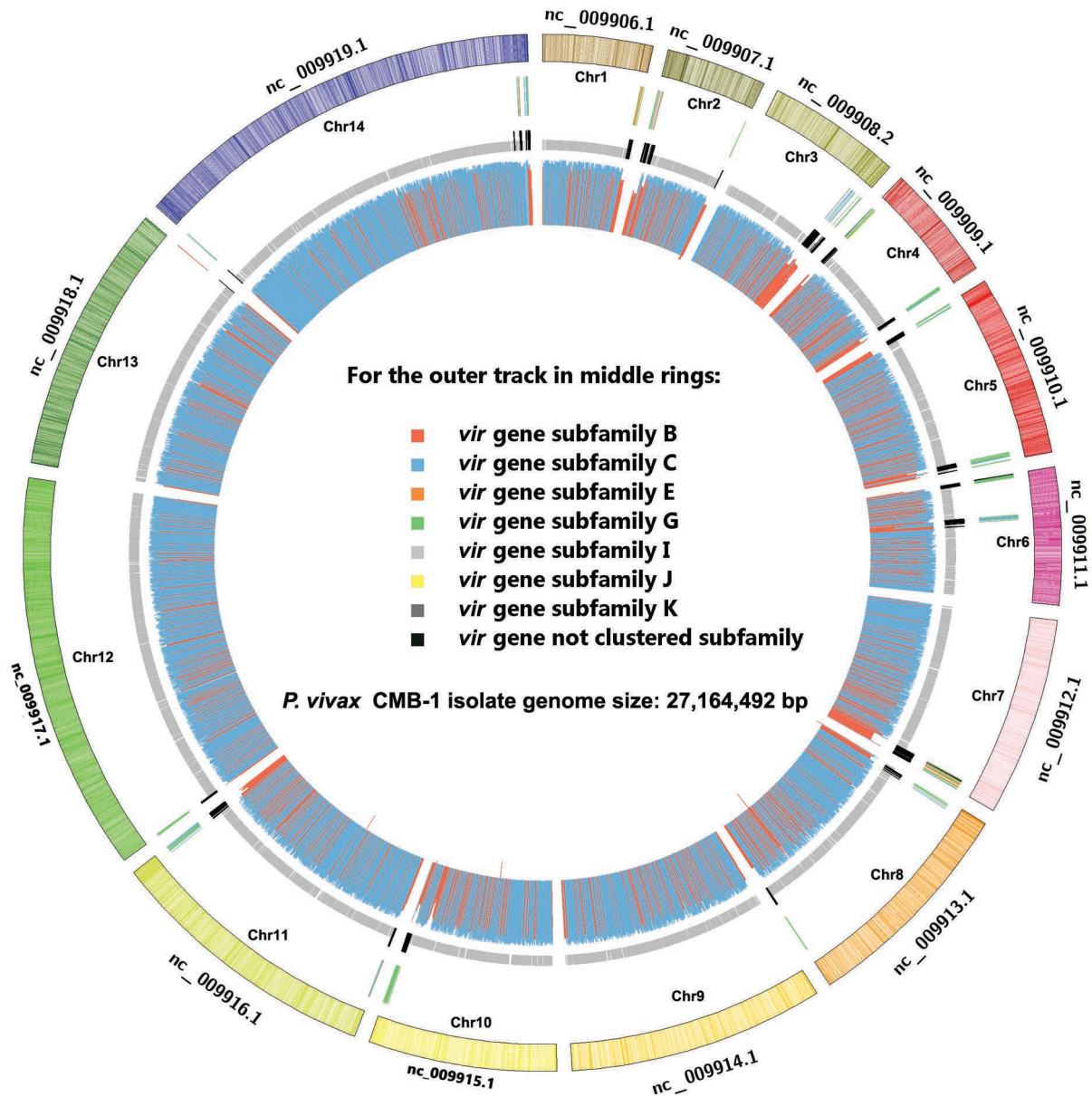
Genomic map of *Plasmodium vivax* China-Myanmar border area (CMB)-1 isolate. The outermost circle shows the whole genome of *P. vivax* Salvador I reference strain. The next ring represents the location of 295 *vir* genes, each subfamily marked in different colour. The third ring shows the coding sequences zone of CMB-1 isolate scaffolds and v*ir* gene orthologs are indicated by black. The inner circle shows the genomic map of *P. vivax* CMB-1 isolate and the histogram represents the degree of similarity (blue: identity > 99%). The figure was drawn using Circos (Krzywinski et al. 2009) and alignment was performed using MUMmer 3.0 (Kurtz et al. 2004).

momatic v.0.30 (Bolger et al. 2014). The sequence reads were aligned to the *P. vivax* Salvador I reference strain (Sal I) genome using BWA-0.7.1 (Li & Durbin 2009). In total, 5.86% of 26 million quality-evaluated reads were aligned onto 96.43% of the Sal I genome with 7.84-fold coverage as well as onto 98.32% of the 14 chromosomes of the Sal I strain covering 95.96-99.05% for each chromosome.

The *de novo* assembly yielded a database with 8,541 scaffolds (10,639 contigs) and an average guanine-cytosine content of 39.1%. A total sequence coverage of 10.26-fold produced this assembly with N50 scaffold

lengths of 5.9 kb. A total of 27.1 Mb of sequence was assembled in the CMB-1 scaffolds (Table).

An overall comparative genomic analysis was conducted using the complete genome of the *P. vivax* reference strain Sal I, as shown in Figure. It is widely known that the *vir* super-family is variably expressed and encodes proteins that are exported to the host cell surface to evade the host adaptive immune response (Fernandez-Becerra et al. 2009). As the largest subtelomeric multigene family of malaria parasites, the *vir* super-family consists of seven different subfamilies. In the CMB-1 *de novo* assembled se-

TABLE

De novo assembly statistics of the *Plasmodium vivax*
China-Myanmar border area-1 genome

Attribute	
Raw reads	15,105,614 paired
Unmapped reads to *Homo sapiens*	1,268,041 paired
After quality control	1,267,552 paired
Contigs (n)	10,639
Scaffolds (n)	8,541
Longest scaffold (bp)	125,157
N50 (bp)	5,936
Genome size (bp)	27,164,492
Coverage	10.26

quences, we identified all published 295 *vir* genes (Lopez et al. 2013) based on their sequence similarity in BLASTX.

The findings in this paper provide whole genomic information on the current epidemiological scenario of vivax malaria in the CMB, where the number of *P. vivax* cases imported from Southeast Asia is increasing and accompanied by growing concern. The results of this work contribute to a better understanding of *P. vivax* evolution and provide an informative basis for further study of the population genomics of this parasite.

REFERENCES

Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics 30*: 2114-2120.

Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature 455*: 757-763.

Coil D, Jospin G, Darling AE 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics 31*: 587-589.

Feng J, Xiao H, Zhang L, Yan H, Feng X, Fang W, Xia Z 2015. The *Plasmodium vivax* in China: decreased in local cases but increased imported cases from Southeast Asia and Africa. *Sci Rep 5*: 8847.

Fernandez-Becerra C, Yamamoto MM, Vencio RZ, Lacerda M, Rosanas-Urgell A, del Portillo HA 2009. *Plasmodium vivax* and the importance of the subtelomeric multigene *vir* superfamily. *Trends Parasitol 25*: 44-51.

Gao Q, Beebe N, Cooper R 2004. Molecular identification of the malaria vectors *Anopheles anthropophagus* and *Anopheles sinensis* (Diptera: Culicidae) in central China using polymerase chain reaction and appraisal of their position within the Hyrcanus group. *J Med Entomol 41*: 5-11.

Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperley WH, Kabaria CW, Tatem AJ, Manh BH, Elyazar IR 2010. The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl Trop Dis 4*: e774.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA 2009. Circos: an information aesthetic for comparative genomics. *Genome Res 19*: 1639-1645.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL 2004. Versatile and open software for comparing large genomes. *Genome Biol 5*: R12.

Li H, Durbin R 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics 25*: 1754-1760.

Lopez FJ, Bernabeu M, Fernandez-Becerra C, del Portillo HA 2013. A new computational approach redefines the subtelomeric *vir* superfamily of *Plasmodium vivax*. *BMC Genomics 14*: 8.

Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S, Gujja S, Goldberg JM, Young S, Zeng Q 2012. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat Genet 44*: 1046-1050.

Price RN, Douglas NM, Anstey NM 2009. New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. *Curr Opin Infect Dis 22*: 430-435.

Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM 2007. *Vivax* malaria: neglected and not benign. *Am J Trop Med Hyg 77*: 79-87.

Winzeler EA 2008. Malaria research in the post-genomic era. *Nature 455*: 751-756.

Zhou X, Huang JL, Njuabe MT, Li SG, Chen JH, Zhou XN 2014. A molecular survey of febrile cases in malaria-endemic areas along China-Myanmar border in Yunnan province, People's Republic of China. *Parasite 21*: 27.