Check for updates

RESEARCH ARTICLE

A tiling array-based comparative genomic hybridization approach to predict copy number variations between *Plasmodium falciparum* field isolates from the Indian Sub-continent [version 1; peer review: 1 approved, 1 approved with reservations]

Isha Pandey^{1*}, Ramandeep Kaur^{1*}, Amit Kumar Subudhi¹, P.A Boopathi¹, Raja C. Mugasimangalam¹, Sudha N. Rao², Mohammed Aiyaz², Sanjay Kochar³, Dhanpat Kochar⁴, Ashis Das ¹

¹Department of Biological Sciences, Birla Institute of Technology & Science, Pilani, Rajasthan, India

²Genotypic Technology Pvt. Ltd., Bangalore, Karnataka, India

³Department of Medicine, Medical College, Sardar Patel University, Bikaner, Bangalore, India

⁴Rajasthan University of Health Sciences, Jaipur, Rajasthan, India

* Equal contributors

V1 First published: 09 Jul 2018, 7:1033
<https://doi.org/10.12688/f1000research.14599.1>
 Latest published: 09 Jul 2018, 7:1033
<https://doi.org/10.12688/f1000research.14599.1>

Abstract

Background: There are several techniques to analyse copy number variation in both research and clinical settings, such as whole genome amplification (sWGA), SNP arrays and one of the most commonly used techniques, array based comparative genomic hybridization (aCGH). In the latter, copy number comparison is obtained between differentially labelled target and reference DNAs by measuring ratio of fluorescence intensity of probes indicating loss or gain in the chromosomal region.

Methods: Here we carry out a comparative analysis between two *Plasmodium falciparum* parasite isolates (Pf-isolate-2 and Pf-isolate-1) causing malaria using array CGH. The array contains approximately 418,577, 60mer custom-designed probes with an average probe spacing of 56 bp. The significant major variations (amplifications and deletions) copy number variations (CNV) in Pf-isolate-2 (Pf-2) in comparison with Pf-isolate-1 (Pf-1), are reported.

Results: CNVs have been seen in all the chromosomes in Pf-2, most of the deletions have been seen mostly in sub-telomeric and telomeric regions of the chromosomes that comprises of variant surface antigen family genes. Apart from the subtelomeric regions other parts of the chromosomes have also shown CNVs. Novel variations, like

Open Peer Review

Approval Status

	1	2
version 1
09 Jul 2018	view	view

- Akash Ranjan**, Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India
- Abhisheka Bansal** , Jawaharlal Nehru University, New Delhi, India
- Tejram Sahu**, Johns Hopkins University Bloomberg School of Public Health, Baltimore, USA

Any reports and responses or comments on the article can be found at the end of the article.

continuous amplification of 28kb region (249817-278491) of chromosome-8, which covers for 3 genes two of which codes for conserved *Plasmodium* proteins with unknown function (MAL8P1.139, PF08_0122) and tRNA pseudouridine synthase, putative (PF08_0123). Amplifications in regions harboring genes like GTP cyclohydrolase I (*GCH-1*, PFL1155W) and ribosomal protein, L24, putative (PFL1150C) of chromosome 12 were seen.

Conclusion: Other than known variations reported earlier, some novel variations have also been seen in the chromosomes of Pf-2. This is an experimental case study reporting major amplifications and deletions in Pf-isolate-2 in comparison with Pf-isolate-1 using a tiling array based comparative genomic hybridization approach.

Keywords

Plasmodium falciparum, Clinical isolates, Indian Subcontinent, tiling array, array CGH

Corresponding author: Ashis Das (adas@pilani.bits-pilani.ac.in)

Author roles: **Pandey I:** Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation; **Kaur R:** Conceptualization, Formal Analysis; **Subudhi AK:** Methodology; **Boopathi PA:** Methodology; **Mugasimangalam RC:** Conceptualization, Writing – Review & Editing; **Rao SN:** Conceptualization, Writing – Review & Editing; **Aiyaz M:** Data Curation, Formal Analysis, Methodology, Writing – Review & Editing; **Kochhar S:** Investigation, Methodology, Writing – Review & Editing; **Kochhar D:** Investigation, Methodology, Writing – Review & Editing; **Das A:** Conceptualization, Methodology, Resources, Supervision, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by a grant from the Department of Biotechnology (DBT), of the Ministry of Science and Technology, New Delhi, India [BT/PR7520/BRB/ 10/481/2006] to A.K.D., S.K.K. and D.K.K.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Pandey I *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Pandey I, Kaur R, Subudhi AK *et al.* **A tiling array-based comparative genomic hybridization approach to predict copy number variations between *Plasmodium falciparum* field isolates from the Indian Sub-continent [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2018, 7:1033 <https://doi.org/10.12688/f1000research.14599.1>

First published: 09 Jul 2018, 7:1033 <https://doi.org/10.12688/f1000research.14599.1>

Introduction

Plasmodium falciparum is the most virulent among the 6 *Plasmodium* species that causes malaria in humans. Current efforts to eliminate the disease have not been successful because of the rapid genetic adaptation and natural selection of the parasite. The currently known causes for genome sequence variations in the parasite are single nucleotide polymorphism (SNP), insertions, deletions of short sequences, major amplifications, translocations, inversions, and allelic variations (Cheeseman *et al.*, 2009). Apart from the host factors these parasitic genetic variations may also lead to the disease severity. *P. falciparum* strains have long been established in continuous culture systems and most genomic analysis has naturally utilized the incredible benefit from such techniques. However, analysis of genome variations from culture adapted parasite lines remains intrinsically very different from analysis of *in-vivo* derived parasite nucleic acid material and the latter could have a major impact on understanding of the parasite biology (LeRoux *et al.*, 2009). Environmental pressures have been known to introduce changes in genomes providing organisms an evolutionary advantage for adaptation and CNVs could be one such change that assists the parasite adaptation to the new environment. These CNVs could enable survival of the parasite under drug pressure. An amplification event in the GTP cyclohydrolase has been frequently reported and the high copy number of *GCH-1* has been shown to confer resistance to anti-folates (Heinberg *et al.*, 2013; Jiang *et al.*, 2008; Kidgell *et al.*, 2006; Nair *et al.*, 2008; Ribacke *et al.*, 2007). The major known parasite molecule that influences the severity of the disease are encoded by Variant Surface Antigen family genes that are located in the sub-telomeric part of the chromosome. *P. falciparum* 3D7 genome sequencing revealed the presence of ~60 var genes followed by other variant surface antigen families, rifin/stevor (~190), Pfmc-2TM (~12) and surfin (~10) (Frech & Chen, 2013; Scherf *et al.*, 2008). Most studied var genes encode for *PfEMP-1* (Erythrocyte membrane protein-1) that play an important role in the pathogenesis by periodic switching and immune evasion. Along with the exported family members, these genes play an important role in cytoadhesion by binding to multiple cell receptors, including ICAM-1, CD36, E-selectin, NCAM and CD31 on micro vascular endothelia, which mediates sequestration that in-turn, leads to the severity of the disease. Recently another receptor, Endothelial protein C receptor (EPCR), a brain specific receptor has been identified as a binding partner of *PfEMP-1* (Smith *et al.*, 2013; Wassmer & Gray, 2017). Var genes are divided into 3 major groups A, B and C and 2 intermediate groups UpsB/A and B/C (Lavstsen *et al.*, 2003). The higher transcript level of var group B and var group C family genes have been reported in severe malaria cases in comparison with the uncomplicated malaria cases from India (Subudhi *et al.*, 2015).

There is limited information regarding complete nuclear genome variation for any of the human malaria parasites from field isolates from the Indian subcontinent, however, Tyagi *et al.* in 2014 have reported evolutionary history of *Plasmodium falciparum* by sequence analysis of mitochondrial genome of Indian isolates. Most studies have focused on analysis of sequence variations in single genes from different parts of India (Kumar *et al.*, 2005;

Kumar *et al.*, 2012; Mishra *et al.*, 2015; Rajesh *et al.*, 2008; Zeeshan *et al.*, 2012). These have focused on SNP's for "drug resistance" marker genes of Indian isolates. Garg *et al.*, 2009; Pathak *et al.*, 2014; Sharma *et al.*, 2015).

Genomics technologies like whole genome microarrays and more recently next-generation sequencing have enabled deeper analysis of various types of genomic variations. Using a reference clone/isolate, copy number variations of other isolates can be identified (Carret *et al.*, 2005). An interesting study by Oyola *et al.* in 2016 details of SNP analysis from dried blood spots using Illumina based sequencing approaches.

Array-based technologies have been used to distinguish CNV's in different diseases (Coughlin *et al.*, 2012). Such techniques have been extensively used in cancer research to identify DNA copy number variability to represent structural variations at higher resolution and increased sensitivity (Liu, 2007; Mockler *et al.*, 2005; Przybytkowski *et al.*, 2011).

The objective of this study was to see whether we could at all distinguish differences between the two parasite isolates taken from individual patients showing differing disease severity using aCGH tiling array approach.

The microarray data discussed in this manuscript has been deposited in the NCBI Gene Expression Omnibus (GEO) under the GEO series accession number (GSE77165).

Methods

Collection of blood sample

Venous blood samples were collected (5ml) from *P. falciparum* infected patients admitted to S.P. Medical College, Bikaner, India. Patient's samples were collected on informed consent according to hospital guidelines. Sample collection was approved by the hospital ethical committee (No.F.(Acad) SPMC/2003/2395). Infection with *P. falciparum* was confirmed by microscopy and rapid diagnostic tests (OptiMal test; Diamed AG, Cressier sur Morat, Switzerland, Falcivax test; Zephyr Biomedical System Goa, India) in the hospital. Blood was immediately (within 15 mins) subjected to density gradient-based separation (Histopaque 1077, Sigma Aldrich, USA) to separate the peripheral blood mononuclear cells (PBMCs) from the infected blood samples following manufactures instruction. Both the fractions were subsequently washed twice with phosphate buffered saline (PBS) and lysed using 4 volumes of TRI-reagent and stored at -80°C. Samples were transported in cold chain to BITS, processed and assessed by 18s r RNA gene based multiplex PCR to eliminate any possibility of *P. vivax* co-infection. Details of the PCR primers used for multiplex PCR for the detection of *P. falciparum* and *P. vivax* are presented in Table 1 as cited in the research articles (Pakalapati *et al.*, 2013). PCR conditions used for the amplifications of 18S r RNA gene involved initial denaturation at 93°C for 2 minutes followed by denaturation at 93°C for 1 minute 30 seconds, annealing at 52°C for 2 minutes, extension at 72°C for 3 minutes, after initial denaturation all the reactions were run for 30 cycles. The TRI protocol was used to allow investigations of both RNA and DNA from the precious

Table 1. Details of the PCR primers used for the multiplex PCR for the detection of *Plasmodium falciparum*, and *P. vivax*, PCR conditions used for the amplifications of 18S r RNA gene involved initial denaturation at 93°C for 2 minutes followed by denaturation at 93°C for 1 minute 30 seconds, annealing at 52°C for 2 minutes, extension at 72°C for 3 minutes, after initial denaturation all the reactions were run for 30 cycles.

	Primer sequence (5'-3')	Detection	Orientation
1	ATC AGC TTT TGA TGT TAG GGT ATT	Genus	Forward
2	TAA CAA GGA CTT CCA AGC	<i>P. vivax</i>	Reverse
3	GCT CAA AGA TAC AAA TAT AAG C	<i>P. falciparum</i>	Reverse

clinical samples. (Chomczynski, 1993) Criteria for determining complicated cases were based on World Health Organization guidelines (World malaria report, 2010). For CGH array hybridization, isolates from two patients were considered a) cerebral malaria (CM) whose Glasgow coma scale = 7/14 was considered as a complicated case (PFC) Pf-2 or test, b) uncomplicated case (PFU) Pf- 1.

Genomic DNA extraction

Total DNA was isolated from samples using back extraction buffer following manufacture's protocol (TRI- Reagent, Sigma Aldrich, and USA). Both the DNA samples (Pf-isolate-2 and Pf-isolate-1) were run on an agarose gel and found to be intact. Total DNA purity and concentration were assessed using Nanodrop ND-1000 UV-Visible spectrophotometer (Nanodrop technologies, Rockland, USA). The A260/A280 ratio suggests the absence of proteins and RNA.

Genomic DNA preparation and hybridization

1.5 Micrograms of each DNA sample were restriction digested using *AluI* and *RsaI*. This restricted DNA was random-primer labelled with Cyanine-3 (Cy3) and Cyanine-5-dUTP (Cy5) using Agilent Genomic DNA labeling kit (Part Number: 5190-0453). The labelled DNA was concentrated and quality assessed for yields and specific activity. Pf-2 which is a test sample was labelled with Cy3 and Pf-1 sample taken as a control was labelled with Cy5. 5 micrograms of the labelled samples were hybridized on a Genotypic designed *P. falciparum* 3D7 Custom 2x400K CGH Microarray, (Agilent Microarray Design ID AMADID: 025327). Hybridization of probes was carried out in Agilent Sure hyb chambers at 65°C with rotation for 40 hours. The hybridized slides were washed using Agilent aCGH wash buffer (Part No: 5188-5221/22) and scanned using the Agilent Microarray Scanner G2505C. The Scanned image was processed using Agilent Feature extraction software version 10.7 and probe intensities were normalized using the linear dye normalization method.

Comparative genome hybridization array design

A high resolution custom CGH array was designed with 418,577 probes, representing the entire *P. falciparum* 3D7 reference genome using the e-Array tool SureDesign v 5.0.1 by the Agilent certified microarray facility of Genotypic Technology, Bengaluru, India. Whole genome sequence for *P. falciparum* 3D7 strain was

downloaded from NCBI genome database; the 60mer probes were designed based on optimal GC% and melting temperature using Probe parser V.1.0.exe (Perl program developed by Genotypic Technology, Bangalore, India) Genotypic Probe parser is the perl code developed by Genotypic Technology to design 60mer Probes from the downloaded genome sequence and calculate the GC% and melting temperature (Tm) for the probes. The 60 mer probes designed by the above mentioned program were then subjected to the Agilent eArray tool and the CGH array designed. The probes were designed by tiling over the DNA sequence. The tiling resolution was 56 bp. The probe length has been set as 60 bp which is the optimised length for Agilent comparative genomic hybridization workflow. The probes were designed by calculating GC% and optimum Tm values of (>65 genome sequence and calculate the GC% and melting temperature. The sequence information, including chromosomal location and direction of transcriptional regulation of *P. falciparum* 3D7 were retrieved from Plasmodium Genome Resource. [PlasmoDB.V.6.0] (Aurrecoechea *et al.*, 2009; Gardner *et al.*, 2002).

Array CGH data analysis and aberration filter to detect copy number variation

Image analysis and data normalization was done by applying the linear dye normalization method using the Agilent Feature Extraction Software version 10.7. The Raw data were analysed using the Agilent Genomic Workbench Software version 6.5. A well-accepted ADM-2 algorithm was used (Aberration Detection Method II algorithm) developed by Agilent and included in the Agilent Genomic Workbench for CGH analysis. ADM-2 is a statistical algorithm which computes copy number differences between the sample and reference using an iterative procedure to identify all genomic regions from its expected value of 0 larger than a given threshold (In this case we have taken a minimum of 5 consecutive probes, with an average log ratio of 0.3). At each iteration, the region with the most significant score is reported. ADM- 2 statistical score is computed as the average normalized Cy3/Cy5 log₂ ratios of all probes in the genomic interval multiplied by the square root of the number of these probes. Deletions are reported where the signal intensities of the probes representing the region in the test strain are near background and represents the absence of hybridization. Graphical visualization of the data for the significant aberrations was done using UCSC genome browser (Schneider *et al.*, 2006).

Results

In this study, a custom designed tiling microarray has been used to investigate the overall difference between the two clinical isolates i.e., Pf-isolate-2 (Pf-2) (complicated/cerebral malaria) and Pf-isolate-1 (Pf-1) (uncomplicated malaria). The reference genome 3D7 was used to design the probes. The tiling array used in this study enables large-scale detection of regions showing copy number variations across the whole genome. The CNV's were identified in Pf-2 (complicated) by comparison to Pf-1 (uncomplicated malaria). The well-accepted ADM-2 algorithm was used with filtering criteria of, minimum of 5 probes and the average log fold change of 0.3 for the analysis of CNV. Here we are reporting significant CNV regions seen in Pf-2 in comparison with the Pf-1. In this data, the term amplification and deletion are used to represent copy number of test strain Pf-2 compared to Pf-1. Deletions are reported where the signal intensities of the probes representing the region in the test strain are near background and represent the absence of hybridization.

The size of each chromosome with the number of genes reported here are as per PlasmoDB V.6.0 (Aurrecochea *et al.*, 2009; Gardner *et al.*, 2002). Those regions that are re-annotated in the genome assembly of *P. falciparum* 3D7 and updated in recent version of PlasmoDB i.e PlasmoDB V.34 and are eliminated from the study to rule out any inadvertent errors due to the re-annotation processes. A circos plot of whole genome CNV data of Pf-Isolate-2 has been generated to show the regions in chromosomes exhibiting CNV (Figure 1).

Genome wide gene copy number variation

A total of 687 unique genes was seen to exhibit CNVs in the genome of Pf-isolate-2 (Dataset 1, (Pandey *et al.*, 2018a)). The loss/gains varied in size from 100 nucleotides to 47,000 bp and included intra and intergenic regions, as they are an essential part of the genome and might be involved in the regulation of the genes. From the total of 687 genes, amplification is reported in 511 and deletions in 176 genes. There are 27 genes in regions of chr-4,6,7,8,10,11,12,13 in Pf-2 in which both amplifications/gains and deletions/loss are present. This region is rich in Variant Surface Antigen family genes. The deletions or amplifications covered smaller regions and did not cover more than 2 contiguous genes. Most variations were seen in the sub-telomeric and telomeric regions.

Variations in Variant Surface Antigens (VSA)

Our analysis revealed CNVs in 268 unique VSA gene families, (Dataset 2, (Pandey *et al.*, 2018b)) out of which variation is present in 58 unique var genes in Pf-2 in comparison to Pf-1. Most of these genes showing CNVs belong to var group B and C. A previous report from our group has also reported differential expression in the transcriptome of these family genes in the isolates causing severe malaria. The longest CNV was reported in chr2: 86, 3297: 94, 7019, near the telomere covering about 20 genes. This region largely had deletions and covered genes belonging to the: var gene family (*PfEMP-1*) rifin, stevor, and exported family members Plasmodium exported protein (*EPF-1*, *EPF-3*, *EPF-4*) Pfmc-2TM Maurer's cleft two

Figure 1. Circos plot of whole genome variation in Pf-isolate-2. Circle of Genome variation of *Plasmodium falciparum* (Pf-2), representing probe based chromosome-wise Copy number variations. The outer red bars are showing amplifications i.e. regions with $Cy3/Cy5$ ratio of ≥ 0.3 and green inner bars are showing deletions present in the region with $Cy3/Cy5$ ratio ≤ -0.3 .

transmembrane protein (*MC-2TM*) and hypothetical proteins. Regions with multigene family genes showed deletions at both the telomeric and centromeric regions of the chromosomes. In chromosome-1 of Pf-isolate-2 a deletion of 16.3kb (563139-579488) region was seen, that includes 5 genes coding for Stevor, pseudogene (PFA0705C) rifin (*RIF*, PFA0710C) Plasmodium exported protein (hyp7), unknown function (PFA0715C) Plasmodium exported protein, unknown function, pseudogene (PFA0720W). There is also a 11.7 kb, deletion region in the same chromosome, which includes 4 genes coding for, Exported protein Family-3 (*EPF-3*, PFA0685C) Exported Protein Family-4 (*EPF-4*, PFA0690W) Erythrocyte Membrane Protein1 exon2, pseudogene (VAR) (*PfEMP-1*, VAR, Pseudogene, PFA0695C), and Plasmodium exported protein (hyp10) of unknown function (*hyp10*, PFA0700C). In previous reports SNPs have been detected in PFA0700C Plasmodium exported protein (hyp10) in malaria parasites from Guinean and Gambian populations (Mobegi *et al.*, 2014). A study by Mok *et al.* in 2008 have shown a segmental duplication of PFA0675W, PFA0685C, PF0690W and PFA0700C in different strains, and another study from the same group have also reported the duplication of PFA0685C, PF0690W and PFA0695C located at the right end of the chromosome of 3D7 in a clinical isolate (UAM25) (Mok *et al.*, 2008; Ribacke *et al.*, 2007). In our clinical isolate showing complicated manifestation, these segments are showing deletion in the chromosome. A deletion of 11.9 kb region in chromosome-3 that includes conserved *Plasmodium* proteins with unknown functions (PFC0325C, PFC0330W, PFC0335C). It is interesting to note that although these genes are of unknown function their annotated GO functions suggest that these genes are involved in zinc ion binding activity, receptor activity and in ATP binding activity. A deletion of 14.8 kb, is present in chromosome number 11 and the region includes 6 genes rifin (*RIF*, PF11_0010), rifin (*RIF*, PF11_0011), RESA-like protein, pseudogene (*RESA* pseudogene, PF11_0012) Plasmodium exported protein (PHISTa), unknown function, pseudogene (*PHISTa*, pseudogene, PF11_0013), Pfmc-2TM Maurer's cleft two transmembrane protein (*MC-2TM*, PF11_0014) and RESA-like protein, pseudogene, (*RESA*, PF11_0015). This deletion region is present at the left end of the chromosome. A size variation of 13.2 kb has been seen in chromosome-13 from a starting position of 2828057-2841275 and covers 5 genes rifin (*RIF*, MAL13P1.495), rifin (*RIF*, MAL13P1.500), Stevor (*Stevor*, MAL13P1.505)

erythrocyte membrane protein, exon-2, pseudogene (VAR) (*PfEMP1*, MAL13P1.510) rifin (*RIF*, MAL13P1.515). Other than larger deletions in chromosomes that cover for VSA, we have seen an amplification of size 13.2 kb present at the right end of chromosome-11 (2006705 - 201923) which covers 5 genes which encodes for rifin (*RIF*, PF11_0515), Stevor (PF11_0516), rifin (*RIF*, PF11_0517), rifin (*RIF*) pseudogene (*RIF*, PF11_0518) and rifin (*RIF*, PF11_0519). Details of the mentioned regions, and other regions covering VSA, are shown in Dataset 3 (Pandey *et al.*, 2018c).

Major variations other than Variant Surface Antigens

On analysis of the CNV data we have detected variations in all the 14 chromosomes; apart from the sub-telomeric regions of the chromosomes, we have also seen variations in other locations on the chromosomes, including intergenic regions.

Other than the regions belonging to multigene family genes, which are mostly showing deletions in the chromosome, there are regions in which amplifications are present and represents genes that encodes for transcriptional co-activator ADA2 (*ADA2*), chromodomain-helicase-DNA-binding protein 1 homolog, putative (*CHD1*), polyubiquitin (*PfpUB*), non-SERCA-type Ca²⁺-transporting P-ATPase (*ATP4*), etc. Major amplifications are present in chromosome-8, 10, and 12. An amplification of 28.6 kb is present in chromosome-8 (249817-278491) (Figure 2), near the centromeric location of the chromosome and covers 3 genes conserved Plasmodium protein, unknown function (MAL8P1139), conserved Plasmodium protein, unknown function (PF08_0122) tRNA pseudouridine synthase, putative (PF08_0123). In chromosome 10, A 7.4 kb and 5.6 kb amplification is present in genes that code for chromodomain helicase DNA binding protein 1 homolog, putative (*CHD1*, PF10_0232) and transcriptional co-activator ADA-2, (*ADA-2*, PF10_0143) Another region from 650889-653243 with an amplification of 2.3 kb cover for Glycophorin Binding Protein (*GBP*, PF10_0159). An amplification of 15.90 kb is present in close proximity to the centromeric region of chromosome 12 (Figure 3), which covers 6 genes that codes for conserved Plasmodium protein with unknown function (PFL1160C), Mitochondrial carrier protein, putative (PFL1145W) conserved plasmodium protein, unknown function (PFL1135C) GTP cyclohydrolase-I *GCH1*, (*GCH-1*, PFL1155W) ribosomal protein L24, putative (PFL1150C)

Figure 2. Continuous amplification of 28 kb region on chromosome-8 of Pf-isolate 2. A continuous amplification of 28 kb region (249817-278491) of chromosome-8, the X-axis represents the baseline intensity of the probes, the Y-axis represents the hybridized probe intensity, red bars are showing amplifications present in the region and green bars are showing deletions present in the region. The genes present in the region, are near centromeric location on the chromosome-8 that covers for 3 genes (MAL8P1.139, PF08_0122) conserved *Plasmodium* membrane protein, unknown function, (PF08_0123) U2 snRNA/tRNA pseudouridine synthase, putative. The white double-headed arrows in single exons are representing the 5th the 5exons are representing the 5epresents the the arrows in between the exons are representing introns and strand in which genes are present (Right-hand side arrow forward strand, left-hand side arrow representing reverse strand). The area without any bars are representing no variations.

Figure 3. Continuous amplification of 15 kb region on chromosome-12 of Pf isolate-2. Amplification of 15.90kb is present near close proximity to the centromere region of chromosome 12 (Figure 3), which covers 6 genes, namely (PFL1160c), conserved *Plasmodium* protein, unknown function, (PFL1145w) mitochondrial carrier protein, putative, (PFL1135c) conserved *Plasmodium* protein, unknown function, (PFL1155w) *GTP cyclohydrolase I (GCH1)*, (PFL1150c) *ribosomal protein L24*, putative, (PFL1140w). The X-axis represents the baseline intensity of the probes, the Y-axis represents the hybridized probe intensity, red bars are representing probe intensity as amplifications present in the region and green bars are showing deletions present in the region. Those without any bar are showing no variations.

and 50S integral membrane protein, putative (PFL1140W) An amplification of the region, that covers three genes, PFL1155W, PFL1150C, and PFL1145W have been seen previously in Dd2 strain of *Plasmodium falciparum* (Dharia *et al.*, 2009). One of the genes showing amplification codes for the GTP cyclohydrolase enzyme, which is the 1st and the rate-limiting enzyme of the folate biosynthesis pathway. Amplification of this gene has been previously reported in parasite isolates from certain endemic countries (Kidgell *et al.*, 2006; Nair *et al.*, 2008). Another gene, mitochondrial carrier protein, putative (PFL1145W) is reported as a homologue of Yhm2 which was characterized from *S. cerevisiae* as an inner membrane DNA binding protein. It is speculated that this protein helps in mitochondrial maintenance by tethering mitochondrial DNA to inner membranes (van Dooren *et al.*, 2006). Another region in chromosome-12 showing amplification covers 2 genes codes for Polyubiquitin (*PfpUB*, PFL0585W) and non-SERCA-type Ca transporting P-ATPase (*ATP-4*, PFL0590C).

An amplification of 3.3 kb is present in chromosome -13 (408857- 412219) in the single gene that codes for mRNA-decapping enzyme 2 putative (*DCP2*, PF13_0048), Amplification of 3.19 present on chromosome-14 covering a single gene coding for pyridine nucleotide transhydrogenase, putative (PF14_0508) and 3.13 kb which is also present in chromosome-14 covering a single gene EMP-1 trafficking protein (*PTP3*, PF14_0758). It has been reported that disruption of the gene (PF14_0758) showed reduced levels of expression of *PfEMP-1*, which suggests that the protein encoded by PF14_0758 plays a role in trafficking and the display of the virulence protein *PfEMP-1* on the host erythrocyte (Maier *et al.*, 2008). All these genes showing amplifications in the data are important for the parasite to survive in the host environment

Dataset 1. Total set of Genes showing Copy Number Variation in Pf-isolate- 2 in comparison with the Pf-isolate-1

<http://dx.doi.org/10.5256/f1000research.14599.d209655>

Includes data of total set of genes showing CNV, along with the Information about the Chromosome Number, start of the probe, end of the probe, size of the interval, number of probes in the interval, Amplification call, Deletion call, Gene symbols (PlasmoDB V.6.0), Updated gene symbol (PlasmoDB V.26), Description of gene symbol (PlasmoDB V. 34) log value of the regions, P-value of the region

Dataset 2. Copy Number Variations present in Variant Surface Antigen family genes of Pf-isolate-2

<http://dx.doi.org/10.5256/f1000research.14599.d209656>

Includes data of the total set of VSA gene family genes that are showing variations. Along with the Information about Chromosome Number, start of the probe, end of the probe, the size of the interval, number of probes in the interval, Amplification call, Deletion call, Gene symbols (PlasmoDB V.6.0), Updated gene symbol (PlasmoDB v. 26), Description of gene symbol (PlasmoDB V. 28) log value of the regions, P-value of the region

Dataset 3. Genes from Variant Surface Antigen (VSA) family genes that are showing more than 2 kb variations

<http://dx.doi.org/10.5256/f1000research.14599.d209657>

Includes data from genes of regions covering for VSA family genes that are showing more than 2 kb variation along with the Information about Chromosome Number, start of the probe, end of the probe, size of the interval, number of probes in the interval, Amplification call, Deletion call, Gene symbols (PlasmoDB V.6.0), Updated gene symbol (PlasmoDB v. 26), Description of gene symbol (PlasmoDB V.34) log value of the regions, p-value of the region

Discussion

Several studies at the genomic and transcriptomic level have been undertaken to understand the molecular basis of pathogenesis involved in severe disease manifestation. However, there is scarcity of information on genomic variation in parasites causing complicated manifestations, regions with seasonal malarial transmission. Besides, most of the CNV or SNP studies have been done on *in-vitro* parasite isolates or clinical isolates from other parts of the world, but very few studies are from parts of India (Cheeseman *et al.*, 2009; Cheeseman *et al.*, 2016). The lacunae in genome variation study still exists as there is no study from parts of India showing nuclear genomic variations between two parasites isolates with differing disease manifestations. CNVs were seen to be present throughout the chromosomes of Pf-2. Large amplifications and deletions of several kilo bases in regions covering *Pf-EMP1* and exporter family member genes, have been seen, as it is a known fact that the variant surface antigen family genes are prone to variations and immune invasions. (Carret *et al.*, 2005; Mok *et al.*, 2008; Tan *et al.*, 2009). Amplification in chromosome number 12 includes genes like *GCH-1* which is a rate limiting enzyme of folate biosynthesis

and is important for the parasite to survive, amplification of this gene is reported to cause antifolate resistance and drives the evolution of drug resistance (Kidgell *et al.*, 2006; Nair *et al.*, 2008). Amplification in Glycophorin binding protein (GBP, PF10_0159) was also seen in the PF-2, this gene is important for the parasite to invade the erythrocyte as it binds to the Glycophorin protein on the erythrocyte membrane. Amplification in sodium-dependent phosphate transporter (PiT, MAL13P1.206) has also been seen, the uptake of Pi is essential for the parasite for the synthesis of DNA, RNA and numerous phosphorylated metabolic intermediates (Saliba *et al.*, 2006).

Conclusion

The implications of deletions or amplifications in the coding regions of the genes remain unanswered at this point. This study shows the presence of CNV based genome differences from two field isolates. Based on msp-1 and msp-2 genotyping it appears that both the isolates exhibit allelic families representing K1 (msp-1 block-2) and FC27 (msp-2 repeats) (Snounou *et al.*, 1999). At this level of the study, it is not possible to speculate whether the differences that we are seeing in Pf-isolate-2 are instrumental in causing the severity of the disease. However, we are reporting here some putative and novel variations, which have not been reported before in single gene studies with field isolates. The Data clearly shows that there is a difference in the isolate which has been taken from the patient with severe malaria (Pf-2) than from that with uncomplicated malaria (Pf-1). Differences exist not only in coding regions, but also in the intergenic locations. Thus, there could be implications at the level of gene regulation. However, these differences could also be inherently natural differences between the two parasite isolates without any direct correlation with the disease states, which could be influenced by various other host factors. This case study is just a preliminary attempt to utilize aCGH tiling array using ADM-2 algorithm to report the variations if any or differences in the genomic segments of parasite isolate causing severe malaria with that of the isolate causing uncomplicated malaria. To get a real picture of the genomic variations related to the disease severity, a tiling array approach can be used to analyze a large set of clinical samples exhibiting different disease complications.

Data availability

The microarray data discussed in this manuscript has been deposited in the NCBI [Gene Expression Omnibus](#) (GEO) under the GEO series accession number ([GSE77165](#)).

Dataset 1: Total set of Genes showing Copy Number Variation in Pf-isolate- 2 in comparison with the Pf-isolate- 1. Includes data of total set of genes showing CNV, along with the Information about the Chromosome Number, start of the probe, end of the probe, size of the interval, number of probes in the interval, Amplification call, Deletion call, Gene symbols (PlasmoDB V.6.0), Updated gene symbol (PlasmoDB V.26), Description of gene

symbol (PlasmoDB V. 34) log value of the regions, P-value of the region. [10.5256/f1000research.14599.d209655](#) (Pandey *et al.*, 2018a)

Dataset 2: Copy Number Variations present in Variant Surface Antigen family genes of Pf-isolate-2. Includes data of the total set of VSA gene family genes that are showing variations. Along with the Information about Chromosome Number, start of the probe, end of the probe, the size of the interval, number of probes in the interval, Amplification call, Deletion call, Gene symbols (PlasmoDB V.6.0), Updated gene symbol (PlasmoDB v. 26), Description of gene symbol (PlasmoDB V. 28) log value of the regions, P-value of the region. [10.5256/f1000research.14599.d209656](#) (Pandey *et al.*, 2018b)

Dataset 3: Genes from Variant Surface Antigen family genes that are showing more than 2 kb variations. Includes data from genes of regions covering for VSA family genes that are showing more than 2 kb variation along with the Information about Chromosome Number, start of the probe, end of the probe, size of the interval, number of probes in the interval, Amplification call, Deletion call, Gene symbols (PlasmoDB V.6.0), Updated gene symbol (PlasmoDB v. 26), Description of gene symbol (PlasmoDB V.34) log value of the regions, *p*-value of the region. [10.5256/f1000research.14599.d209657](#) (Pandey *et al.*, 2018c)

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by a grant from the Department of Biotechnology (DBT), of the Ministry of Science and Technology, New Delhi, India [BT/PR7520/BRB/ 10/481/2006] to A.K.D., S.K.K. and D.K.K.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgement

We thank all the patients and technical workers for their participation and support of this project. I.P. & R.K. acknowledge the Basic Scientific Research fellowship from University Grant Commission, New Delhi, India. A.K.S. acknowledges Senior Research Fellowship from the Council of Scientific and Industrial Research (CSIR), New Delhi, India and Project Assistantship from Department of Biotechnology (DBT), New Delhi, India. P.A.B. also acknowledges earlier Project Assistantship from Department of Biotechnology (DBT), New Delhi, India. A.K.D., S.K.K. and D.K.K. acknowledge and Birla Institute of Technology and Science, Pilani, India and S.P. Medical College, Bikaner, India for providing the required infrastructural facilities during this study.

References

Aurrecoechea C, Brestelli J, Brunk BP, *et al.*: **PlasmoDB: a functional genomic database for malaria parasites.** *Nucleic Acids Res.* 2009; 37(Database issue): D539–D543.

D539–D543.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Carret CK, Horrocks P, Konfortov B, *et al.*: Microarray-based comparative genomic analyses of the human malaria parasite *Plasmodium falciparum* using Affymetrix arrays. *Mol Biochem Parasitol.* 2005; 144(2): 177–186.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cheeseman IH, Gomez-Escobar N, Carret CK, *et al.*: Gene copy number variation throughout the *Plasmodium falciparum* genome. *BMC Genomics.* 2009; 10: 353.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cheeseman IH, Miller B, Tan JC, *et al.*: Population Structure Shapes Copy Number Variation in Malaria Parasites. *Mol Biol Evol.* 2016; 33(3): 603–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chomczynski P: A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques.* 1993; 15(3): 532–4.
[PubMed Abstract](#)
- Coughlin CR 2nd, Schärer GH, Shaikh TH: Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. *Genome Med.* 2012; 4(10): 80.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dharía NV, Sidhu AB, Cassera MB, *et al.*: Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*. *Genome Biol.* 2009; 10(2): R21.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Frech C, Chen N: Variant surface antigens of malaria parasites: functional and evolutionary insights from comparative gene family classification and analysis. *BMC Genomics.* 2013; 14(1): 427.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gardner MJ, Hall N, Fung E, *et al.*: Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature.* 2002; 419(6906): 498–511.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Garg S, Saxena V, Kanchan S, *et al.*: Novel point mutations in sulfadoxine resistance genes of *Plasmodium falciparum* from India. *Acta Trop.* 2009; 110(1): 75–79.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Heinberg A, Siu E, Stern C, *et al.*: Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in *Plasmodium falciparum*. *Mol Microbiol.* 2013; 88(4): 702–712.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jiang H, Yi M, Mu J, *et al.*: Detection of genome-wide polymorphisms in the AT-rich *Plasmodium falciparum* genome using a high-density microarray. *BMC Genomics.* 2008; 9: 398.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kidgell C, Volkman SK, Daily J, *et al.*: A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog.* 2006; 2(6): e57.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kochar DK, Saxena V, Singh N, *et al.*: *Plasmodium vivax* malaria. *Emerg Infect Dis.* 2005; 11(1): 132–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kumar N, Singh JP, Pande V, *et al.*: Genetic variation in histidine rich proteins among Indian *Plasmodium falciparum* population: possible cause of variable sensitivity of malaria rapid diagnostic tests. *Malar J.* 2012; 11(1): 298.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kumar SV, Ranjan S, Saxena V, *et al.*: *Plasmodium falciparum*: genetic diversity of C-terminal region of MSP-1 in isolates from Indian sub-continent. *Exp Parasitol.* 2005; 110(4): 384–388.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lavstsen T, Salanti A, Jensen AT, *et al.*: Sub-grouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar J.* 2003; 2(1): 27.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- LeRoux M, Lakshmanan V, Daily JP: *Plasmodium falciparum* biology: analysis of *in vitro* versus *in vivo* growth conditions. *Trends Parasitol.* 2009; 25(10): 474–481.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Liu XS: Getting started in tiling microarray analysis. *PLoS Comput Biol.* 2007; 3(10): 1842–4.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Maier AG, Rug M, O'Neill MT, *et al.*: Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. *Cell.* 2008; 134(1): 48–61.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mishra N, Prajapati SK, Kaitholia K, *et al.*: Surveillance of artemisinin resistance in *Plasmodium falciparum* in India using the kelch13 molecular marker. *Antimicrob Agents Chemother.* 2015; 59(5): 2548–2553.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mobegi VA, Duffy CW, Amambua-Ngwa A, *et al.*: Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol Biol Evol.* 2014; 31(6): 1490–1499.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mockler TC, Chan S, Sundaresan A, *et al.*: Applications of DNA tiling arrays for whole-genome analysis. *Genomics.* 2005; 85(1): 1–15.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mok BW, Ribacke U, Sherwood E, *et al.*: A highly conserved segmental duplication in the subtelomeres of *Plasmodium falciparum* chromosomes varies in copy number. *Malar J.* 2008; 7(1): 46.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nair S, Miller B, Barends M, *et al.*: Adaptive copy number evolution in malaria parasites. *PLoS Genet.* 2008; 4(10): e1000243.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oyola SO, Ariani CV, Hamilton WL, *et al.*: Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malar J.* 2016; 15(1): 597.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pakalapati D, Garg S, Middha S, *et al.*: Comparative evaluation of microscopy, OptiMAL(®) and 18S rRNA gene based multiplex PCR for detection of *Plasmodium falciparum* & *Plasmodium vivax* from field isolates of Bikaner, India. *Asian Pac J Trop Med.* 2013; 6(5): 346–351.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Pandey I, Kaur R, Subudhi AK, *et al.*: Dataset 1 in: A tiling array-based comparative genomic hybridization approach to predict copy number variations between *Plasmodium falciparum* field isolates from the Indian Sub-continent. *F1000Research.* 2018a.
[Data Source](#)
- Pandey I, Kaur R, Subudhi AK, *et al.*: Dataset 2 in: A tiling array-based comparative genomic hybridization approach to predict copy number variations between *Plasmodium falciparum* field isolates from the Indian Sub-continent. *F1000Research.* 2018b.
[Data Source](#)
- Pandey I, Kaur R, Subudhi AK, *et al.*: Dataset 3 in: A tiling array-based comparative genomic hybridization approach to predict copy number variations between *Plasmodium falciparum* field isolates from the Indian Sub-continent. *F1000Research.* 2018c.
[Data Source](#)
- Pathak A, Mårtensson A, Gawariker S, *et al.*: Characterization of drug resistance associated genetic polymorphisms among *Plasmodium falciparum* field isolates in Ujjain, Madhya Pradesh, India. *Malar J.* 2014; 13(1): 182.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Przybytkowski E, Ferrario C, Basik M: The use of ultra-dense array CGH analysis for the discovery of micro-copy number alterations and gene fusions in the cancer genome. *BMC Med Genomics.* 2011; 4(1): 16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rajesh V, Singamsetti VK, Vidya S, *et al.*: *Plasmodium falciparum*: genetic polymorphism in apical membrane antigen-1 gene from Indian isolates. *Exp Parasitol.* 2008; 119(1): 144–151.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ribacke U, Mok BW, Wirta V, *et al.*: Genome wide gene amplifications and deletions in *Plasmodium falciparum*. *Mol Biochem Parasitol.* 2007; 155(1): 33–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Saliba KJ, Martin RE, Bröer A, *et al.*: Sodium-dependent uptake of inorganic phosphate by the intracellular malaria parasite. *Nature.* 2006; 443(7111): 582–585.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Scherf A, Lopez-Rubio JJ, Riviere L: Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol.* 2008; 62: 445–470.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schneider KL: The UCSC Archaeal Genome Browser. *Nucleic Acids Res.* 2006; 34(Database issue): D407–D410.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sharma D, Lather M, Mallick PK, *et al.*: Polymorphism in drug resistance genes dihydrofolate reductase and dihydropteroate synthase in *Plasmodium falciparum* in some states of India. *Parasit Vectors.* 2015; 8(1): 471.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Smith JD, Rowe JA, Higgins MK, *et al.*: Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. *Cell Microbiol.* 2013; 15(12): 1976–1983.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shounou G, Zhu X, Siripoon N, *et al.*: Biased distribution of *msp1* and *msp2* allelic variants in *Plasmodium falciparum* populations in Thailand. *Trans R Soc Trop Med Hyg.* 1999; 93(4): 369–374.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Subudhi AK, Boopathi PA, Garg S, *et al.*: Natural antisense transcripts in *Plasmodium falciparum* isolates from patients with complicated malaria. *Exp Parasitol.* 2014; 141: 39–54.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Subudhi AK, Boopathi PA, Pandey I, *et al.*: Disease specific modules and hub genes for intervention strategies: A co-expression network based approach for *Plasmodium falciparum* clinical isolates. *Infect Genet Evol.* 2015; 35: 96–108.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tan JC, Patel JJ, Tan A, *et al.*: Optimizing comparative genomic hybridization probes for genotyping and SNP detection in *Plasmodium falciparum*. *Genomics.* 2009; 93(6): 543–50.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tyagi S, Pande V, Das A: Mitochondrial genome sequence diversity of Indian *Plasmodium falciparum* isolates. *Mem Inst Oswaldo Cruz.* 2014; 109(4): 494–498.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

van Dooren GG, Stimmler LM, McFadden GI: **Metabolic maps and functions of the *Plasmodium* mitochondrion.** *FEMS Microbiol Rev.* 2006; **30**(4): 596–630.

[PubMed Abstract](#) | [Publisher Full Text](#)

Wassmer SC, Grau GE: **Severe malaria: what's new on the pathogenesis front?** *Int J Parasitol.* 2017; **47**(2–3): 145–152.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

World malaria report 2010. Geneva: World Health Organization, 2010.

Reference Source

Zeeshan M, Alam MT, Vinayak S, *et al.*: **Genetic variation in the *Plasmodium falciparum* circumsporozoite protein in India and its relevance to RTS,S malaria vaccine.** *PLoS One.* 2012; **7**(8): e43430.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 12 October 2018

<https://doi.org/10.5256/f1000research.15888.r38678>

© 2018 Bansal A et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abhisheka Bansal

School of Life Sciences, Jawaharlal Nehru University, New Delhi, Delhi, India

Tejram Sahu

Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

In the present manuscript by Pandey et al, 'A tiling array-based.... Indian Sub-continent', the authors have used an array based comparative genome hybridization technique (aCGH) to identify genomic variations including copy number variations, amplifications, and deletions in field isolates from Indian sub-continent. The authors have used the technique to compare two *Plasmodium falciparum* field isolates with the objective of identifying genomic differences in Pf-isolate-2 (*Plasmodium* causing complicated malaria) and Pf-isolate-1 (uncomplicated malaria). The study is largely a case study reporting the genomic differences in Pf-isolate-2. The study has identified amplifications in all the chromosomes of Pf-isolate-2 that are largely limited to telomeric, and sub-telomeric regions and include multigene families such as *vars*, *rifins*, and *stevors*. Though, this is not the first report of aCGH technique being used for comparative genome analysis, it's application in study of field isolates is demonstrated. The study holds merit for publication in the journal however, there are some criticisms that need to be addressed.

Major points:

1. The variations including amplifications and deletions are of common occurrence between two parasite strains isolated from different patients. These variations are more prominent in telomeric and sub-telomeric regions comprising of multigene families such as *var* genes, etc. Even two different isolates exhibiting identical phenotypic expression may show variations in multigene families. Since the present study is focused on identifying the genomic differences in Pf-isolate-2 (causing complicated malaria) with Pf-isolate-1 (causing uncomplicated malaria), there is no experimental evidence suggestive of direct association of the CNVs found in the study with the disease phenotype. Moreover, it would have been beneficial to incorporate more samples for comparison including parasites exhibiting identical phenotypes. This would have been useful in drawing meaningful conclusions on the CNVs and disease severity.
2. What variations in results are expected if the study were to use Next Generation Sequencing

- (NSG) instead of aCGH, a summary in the discussion section will be appropriate?
3. How the results might be affected if a newer version of *Plasmodium* genome database was used?
 4. Mention of each gene name in the text, to describe changes, is confusing and it can be presented in the table format making it easier to read and understand.
 5. There are genes such as *msp2* (Chaorattanakawee et al, 2018; PMID: 29342212), etc that have been demonstrated to be associated with disease severity. Did the authors found any variation in *msp2*? The findings of such studies reporting genetic markers of disease severity should be discussed in detail in the discussion section.

Minor points:

Page 3, Introduction section: references for studies that have identified the receptors for EMP1s should be given such as ICAM-1, CD36, E-selectin, NCAM and CD31

Page 3 paragraph 3 last sentence is not necessary, just include the reference.

Page 3 paragraph 5: Objective statement is too long. Reframe it to something like "The objective of this study was to identify genetic difference(s) between two parasite isolates, with varying disease severity, using aCGH tilling approach"

Page 4 Table 1 legend: No need to mention the PCR condition here as it has already been mentioned in method section in Page 3 paragraph 7.

Page 4 paragraph 5: Please mention what is the "GC%" and "melting temperature" considered for designing 60mer probe.

Page 4 paragraph 5: Reframe the sentence "The probes were designed by calculating GC% and optimum Tm values of (>65 genome sequence and calculate the GC% and melting temperature." The *Plasmodium* Genome Resource used was very old (V 6.0) need to be verified with the current version of Plasmo DB (V 39.0).

Please be consistent with *Plasmodium* Gene nomenclature as the Gene ID used were from different version of Plasmo DB it makes it difficult to follow.

Grammatical mistakes: whole manuscript needs a relook to correct the grammatical and typographical errors.

Page 6 Figure 2 legend: Change the font of the sentence "conserved Plasmodium membrane protein, unknown function, (PF08_0123) U2 snRNA/tRNA pseudouridine synthase, putative" from italics to normal. Moreover, the legend has many typographical mistakes, for example: "5the", "5exones", "5epresents", and "the the".

Page 7 Figure 3 legend: Similar to Figure 2 legend it has also few typographical errors that needs to be corrected

Page 7 All the boxes: the "V" in PlasmoDB Version as well as "P" of "P-value" needs to be consistent.

Page 7 last paragraph: Rephrase the sentence "However, there is scarcity of information on genomic variation in parasites causing complicated manifestations, regions with seasonal malarial transmission."

Page 7 last paragraph line 16: Replace "a known fact" with "established"

Page 7 last paragraph line 18: replace "invasion" with "evasion"

Page 8 paragraph 1 line 5: "Pf-2" and not "PF-2"

Reference for datasets presented in the manuscript is not required. It is an integral part of the manuscript.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Reviewer Report 14 September 2018

<https://doi.org/10.5256/f1000research.15888.r37246>

© 2018 Ranjan A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Akash Ranjan

Computational and Functional Genomics Group, Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, Telangana, India

Authors report an aCGH based comparative genomic hybridization approach/case report to detect copy number variations (CNVs) between genes/genomic locus of *Plasmodium falciparum* field isolates (Pf-isolate-1 and Pf-isolate-2) from the Indian Sub-continent. This is an experimental case study reporting major amplifications and deletions in Pf-isolate-2 in comparison with Pf-isolate-1 using a tiling array-based comparative genomic hybridization approach. Authors have used this approach to get a real picture of the genomic variations between the two isolates related to the disease severity.

Authors have identified CNVs in all the chromosomes in Pf-isolate-2 with most of the deletions have been seen in sub-telomeric and telomeric regions of the chromosomes that comprise off variant surface antigen family genes.

Apart from the subtelomeric regions, other parts of *P. falciparum* chromosomes have also shown CNVs. Authors have identified some novel variations, like an amplification of 28kb region (249817-278491) of chromosome-8, which covers for three genes two of which codes for conserved Plasmodium proteins with unknown function (MAL8P1.139, PF08_0122) and tRNA pseudouridine synthase, putative (PF08_0123). Also, amplifications in regions harboring genes like GTP cyclohydrolase I (GCH-1, PFL1155W) and ribosomal protein, L24, putative (PFL1150C) of chromosome 12 were seen.

While authors have successfully used their approach to demonstrate the CNVs between the Pf-isolate-2 (an isolates from severe malaria case) and the Pf-isolate-1 (an isolate without severe malaria complication) very little efforts were made to understand this difference in terms of how these CNVs actually explain the disease severity case. This make the manuscript observational/case study type in nature. Nevertheless, this manuscript has provided a glimpse of *P. falciparum* chromosomal changes that may be associated or responsible for disease severity associated with *P. falciparum* infection in the field. It has potential to stimulate focused investigations to understand the role of specific amplifications and deletions in severity *P. falciparum* infection cases.

Minor changes suggested

1. Page 1: Please verify the affiliation mark for all the authors there could be a mistake in this, specifically the mark for Raja C. Mugasimangalam.
2. English usage/writing eg articles, punctuations, hyphenation, spelling, capitalization, subject verb agreement etc:

Page 3 paragraph 1 line 14: "culture-adapted" and not "culture adapted"

Page 3 paragraph 1 line 2: "that cause" and not "that causes "

Page 3 paragraph 1 line 17: "the understanding" and not "understanding"

Page 3 paragraph 1 line 19: "with an evolutionary advantage" and not "an evolutionary advantage"

Page 3 paragraph 1 line 40: "microvascular" and not "micro vascular"

Page 3 paragraph 2 line 4: "an evolutionary history" and not "evolutionary history"

Page 3 paragraph 2 line 5: "of the mitochondrial genome" and not "of mitochondrial genome"

Page 3 paragraph 2 line 6: "on the analysis of sequence variations" and not "on analysis of sequence variations"

Page 3 paragraph 7 line 17: "gene-based" and not "gene based"

Page 3 paragraph 7 line 22: "amplification" and not "amplifications"

Page 4 Table Title: "amplification" and not "amplifications"

Page 4 paragraph 1 line 2: "the World Health Organization" and not "World Health Organization"

Page 4 paragraph 2 line 2: "the manufacturer's" and not "manufacture's"

Page 4 paragraph 2 line 6: "Nanodrop Technologies" and not "Nanodrop technologies"

Page 4 paragraph 3 line 1: "restriction-digested" and not "restriction digested"

Page 4 paragraph 3 line 4: "Agilent Genomic DNA Labeling Kit" and not "Agilent Genomic DNA labeling kit "

Page 4 paragraph 3 line 12: "Agilent SureHyb chambers" and not "Agilent Sure hyb chambers"

Page 4 paragraph 3 line 13: "Agilent aCGH Wash Buffer" and not "Agilent aCGH wash buffer"

Page 4 paragraph 4 line 1: "high-resolution" and not "high resolution"

Page 4 paragraph 4 line 10: "the Perl code" and not "the perl code"

Page 4 paragraph 4 line 13: "above-mentioned" and not "above mentioned"

Page 4 paragraph 4 line 23: "the Plasmodium Genome Resource" and not "Plasmodium Genome

Resource"

Page 4 paragraph 5 line 19: "and represent" and not "and represents"

Page 4 paragraph 5 line 21: "the UCSC genome browser" and not "UCSC genome browser"

Page 5 paragraph 2 line 4: "in the recent version" and not "in recent version "

Page 5 paragraph 2 line 7: "A Circos plot" and not "A circos plot"

Page 5 paragraph 3 heading: "Genome-wide" and not "Genome wide"

Page 5 paragraph 3 line 4: "intra- and intergenic regions" and not "intra and intergenic regions"

Page 6 paragraph 1 Line 9: "an 11.7 kb," and not "a 11.7 kb,"

Page 6 paragraph 1 Line 15: "In previous reports, " and not "In previous reports "

Page 6 paragraph 1 Line 31: "in the zinc-ion binding" and not "in zinc ion binding "

Page 6 paragraph 1 Line 38: "transmembrane" and not "transmembrane"

Page 6 paragraph 1 last but one line: "other regions covering VSA," and not "other regions covering VSA "

Page 6 Fig 2 legend: "The genes present in the region are near centromere on the chromosome-8" and not "The genes present in the region, are near centromeric location on the chromosome-8"

Page 6 Fig 2 legend: "the arrows" and not "the the arrows"

Page 6 Fig 2 legend: "The white double-headed arrows in single exons are representing the 5th exons are representing the 5th presents the the arrows in between the exons are representing introns and strand in which genes are present (Right-hand side arrow forward strand, left-hand side arrow representing reverse strand)." Not clear please rewrite.

Page 7 paragraph 1 Line 1: "the 50S integral membrane protein" and not "50S integral memberane protein"

Page 7 paragraph 1 Line 4: "the Dd2" and not "Dd2"

Page 7 paragraph 2 Line 5: "nucleotide transhydrogenase" and not "nucelotide transhydrogeanse"

Page 7 title: "The set of total genes" and not "Total set of Genes"

Page 7 paragraph 3 Line 4: "a scarcity of information" and not "scarcity of information"

Page 7 paragraph 3 Line 5: "in the regions with seasonal malaria transmission" and not ", regions with seasonal malarial transmission"

Page 7 paragraph 3 Line 9: "A lacuna in genome variation study still exists" and not "The lacunae in genome variation study still exists"

Page 7 paragraph 3 Line 14: "kilobases" and not "kilo bases"

Page 7 paragraph 3 Line 15: "genes have been seen" and not "genes, have been seen"

Page 7 paragraph 3 Last Line : "rate-limiting" and not "rate limiting"

Page 8 paragraph 2 Line 15: "the coding regions" and not "coding regions,"

Page 8 paragraph 2 Line 21: "the ADM-2" and not "ADM-2"

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research