Journal of Communicable Diseases
Volume 51, Issue 4 - 2019, Pg. No. 29-40
Peer Reviewed & Open Access Journal

Research Article

Genomic Glimpse of the Chromatin Modifier SET Domain family in *Plasmodium falciparum*

Manjeri Kaushik, Priyanka Chahar, Ashima Nehra, Naresh Kumar, Ritu Gill

Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India.
DOI: https://doi.org/10.24321/0019.5138.201934

I N F O

Corresponding Author:
Ritu Gill, Centre for Biotechnology, Maharshi Dayanand University, Rohtak, Haryana, India.
E-mail Id:
rgillcbt@gmail.com
Orcid Id:
https://orcid.org/0000-0002-8959-3089
How to cite this article:
Kaushik M, Chahar P, Nehra A, Kumar N, Gill R. Genomic Glimpse of the Chromatin Modifier SET Domain family in *Plasmodium falciparum*. *J Commun Dis* 2019; 51(4): 29-40.

Date of Submission: 2019-12-15
Date of Acceptance: 2019-01-03

A B S T R A C T

Histones N-terminal tails are the sites for Post-Translational Modifications (PTMs) that regulate the chromatin structure, thus chromatin associated processes. PTMs include methylation, acetylation, phosphorylation, ubiquitination, sumoylation, and ribosylation. Histone lysine methylation is associated with both transcription activation and repression. The SET domain proteins carry out the histone lysine methylation on the N-terminal tails of histones H3 and H4 and are called Histone Lysine Methyltransferases (HKMTs). A total of ten SET domain genes have been identified in human malarial parasite *Plasmodium falciparum*. The present study provides detailed computational analysis of *P. falciparum* SET domain proteins (*Pf*SETs). The analyses cover *Pf*SET family in terms of domain composition, physiochemical properties, subcellular localization, expression profiling and phylogenetic relationships. The work also highlights the conservation of important catalytic residues in *Pf*SETs. The present study provides a detailed insight into the *Pf*SET family, thus opens a platform for further developments.

Keywords: Malaria, *Plasmodium falciparum*, SET domain, Histone Lysine Methyltransferases

Introduction

Despite significant reduction in malaria incidences since 2010, malaria continues to be a major health concern affecting millions of people in tropical and subtropical regions.[1] Malaria parasite life cycle occurs in two different hosts- a vertebrate and an invertebrate. The pathogen exhibits differential gene expression to cope up with distinct environments it experiences during its life cycle.[2-5] However, regulation of gene expression is hugely understudied in malaria parasite. Histones covalent modifications (methylation, acetylation, phosphorylation, ubiquitination, sumoylation, ribosylation etc.) carried by chromatin modifying proteins play a significant role in regulation of chromatin structure and gene expression.[6] Chromatin mediated epigenetic regulation plays an important role in malaria parasite.[7] Histone methylation is a widespread covalent modification that occurs on lysine and arginine residues in the N-terminal tails of histones H3 and H4.[8] Histone lysine methylation is involved in a number of processes such as transcriptional regulation, heterochromatin formation, DNA damage response and cancer.[9,10]

Lysine methylation on the histone tails is carried out by a group of proteins containing SET domain. SET domain is approximately 130 amino acids long evolutionary conserved motif present in chromatin associated proteins from yeast to mammals.[11,12] SET domain proteins act as histone lysine methyl transferases (HKMTs) that transfer methyl group from the cofactor S-adenosyl-L-methione (SAM) to lysine residues of the histone tails. PKMTs regulate

transcription and other cellular functions through site-specific methylation of histones and other substrates.[13] The SET domain is named after its first identification in three *Drosophila melanogaster* proteins: Su(var)3-9; Suppressor of variegation 3-9, E(z); Enhancer of zeste and Trx; the trithorax-group.[14-16]

The differential gene expression patterns in different stages of *Plasmodium* life cycle are maintained by number of mechanisms, the most important are covalent histone modifications and three-dimensional chromatin conformation.[17, 18] Methylation marks in H3 and H4 histones are involved in regulation of the parasite erythrocytic cycle especially schizont stage.[19] *Pf*SET vs directed methylation of histone H3K36 results in the repression of all var genes and allows expression of one gene at a time. The mechanism of expression of only one out of the 60 var genes leads to the antigenic variation of the parasite and thus evasion of the host immune system.[19]

As SET proteins are involved in various vital cellular and biochemical functions, thus play important role in progression of *P. falciparum* through different developmental stages. Further, targeting these histone methylation writers *Pf*SETs could be an important strategy in controlling the intractable parasite. Mining of SET genes from *P. falciparum* and their further bioinformatics analysis will extend our understanding of these proteins in *P. falciparum* life cycle that can be potential drug targets in future.

Materials and Methods

Identification of SET Domain Containing Genes in *P. falciparum* (*Pf*SET genes)

Plasmodium genomic resource PlasmoDB[20] version 9.0 (http://PlasmoDB.org) was searched to identify *P. falciparum* SET domain encoding genes (*Pf*SET genes). Gene text search was primarily used to collect putative *Pf*SET genes from PlasmoDB. Extraction of all *Pf*SET genes was ensured by BLASTp analysis carried out with a threshold expect value of ≤10 using protein sequences of SET domain from various organisms (*Drosophila melanogaster, Arabidopsis thaliana, Homo sapiens)* as a query and low complexity filter turned off.

Confirmation of SET Domain in Putative Genes

To validate the presence of SET domain in proteins encoded by putative *Pf*SET genes, SMART[21] (http://smart.embl.de/), InterPro (http://www.ebi.ac.uk/interpro/) databases were explored. Protein sequences of each putative *Pf*SET genes were obtained from PlasmoDB followed by confirmation of predicted SET domain by SMART/InterPro.

Analysis of *Pf*SET Genes

All the information about *Pf*SET genes regarding Gene Ids, chromosomal location, genomic position, and number of introns, nucleotide sequence length, molecular weight, amino acid sequence length and Isoelectric Point (IP) was extracted from PlasmoDB. Number of SET domains in a gene was identified by SMART.

Domain Architecture of *Pf*SET Proteins

Domain structures of all *Pf*SET proteins were identified with the SMART database. All the protein sequences of *Pf*SETs were submitted to the SMART database one by one. Domain architecture of all *Pf*SET proteins was drawn manually. Domains were named as identified by SMART.

Prediction of Subcellular Localization of SET Proteins

Various online softwares like Mitoprot[22] (http://ihg.gsf.de/ihg/mitoprot.html), Euk-mPLoc 2.0 server[23] and PATS[24] were used to predict subcellular localization of *Pf*SET proteins in *P.falciparum*. Mitoprot (http://ihg.gsf.de/ihg/mitoprot.html) is a web based computer program for predicting mitochondrion proteins. Any sequence showing mitochondria targeting probability equal to or greater than 0.4 was considered to be a mitochondrial protein. Euk-mPLoc 2.0 server (http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/) is used for predicting sub-cellular localization of eukaryotic proteins. PATS (http://modlabcadd.ethz.ch/software/pats/) was used to identify the apicoplast targeted *Pf*SETs.

Expression Profiling

Expression of all *Pf*SET genes at different stages of life cycle of P. falciparum was analysed as per available transcriptome and proteome data at PlasmoDB. Proteomic data was retrieved from Florens et al.[2], Lasonder et al.[3], Le Roch et al.[5], Silvestrini et al.[25], Solyakov et al.[26], Treeck et al.[27], Oehring et al.[28], Lindner et al.[29], Pease et al.[30] Transcriptome data of intraerythrocytic stages as provided by DeRisi group[31] was analysed for existence of mRNA of *Pf*SET genes. The heat maps for *Pf*SETs were constructed using MeV software version 4.9.

Prediction of Human Homologs of *Pf*SET Proteins

Homologs of *Pf*SET proteins in *Homo sapiens* were predicted by BLASTp analysis of individual *Pf*SET protein sequences against HPRD (Human Protein Reference Database). HPRD[32] is freely available at www.hprd.org. The best hit was taken and further analysed to validate it as a homolog by its domain architecture (SMART and Pfam), functional annotation and protein size. Homologs list was prepared showing NCBI gene ids and product description against the respective *Pf*SET gene.

Multiple Alignment Sequence and Identification of Conserved Features

Protein sequences of only SET domains for all *Pf*SETs were

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

extracted as identified by SMART database. All the extracted SET domain sequences were aligned with CLUSTAL Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) using default parameters. SET specific conserved residues/signature motifs were identified based on multiple sequence alignment and highlighted manually.

Generation of Sequence Logo

To represent level of conservation of specific residues at particular positions in the *Pf*SET domains, sequence logo for *Pf*SET domain was generated using online software WebLogo (http://weblogo.berkeley.edu).

Formation of Phylogenetic Tree

The evolutionary connections between organisms are represented graphically through phylogenetic trees. Multiple sequence alignment of protein sequences was performed using the ClustalW program. The resulting file was subjected to phylogenetic analysis using the program PHYLogeny Inference Package (PHYLIP version 3.69).[33] An un-rooted neighbor-joining Phylogenetic tree was constructed by generating 100 random bootstrap replicates using PHYLIP. Resulting tree file was visualized by Mega 5.05 program.

Results and Discussion

Identification and Listing of *Pf*SET Genes in *P. falciparum*

By gene text searching and BLASTp analysis carried out with a threshold e-value of ≤10 using protein sequences of SET domain from various organisms (*D. melanogaster, A. thaliana, H. sapiens*) as a query, we were able to extract total 10 putative *Pf*SET genes from PlasmoDB. However, out of 10 putative *Pf*SET genes, SET domain was confirmed in 8 *Pf*SET genes by Pfam and SMART databases (Table 1). Remaining 2 putative *Pf*SET genes did not encode any SET domain according to Pfam and SMART. The HKMTs without SET domain are reported.[34] A previous study although reported nine SETgenes in *P. falciparum*.[35]

Analysis of *Pf*SET Genes

Total 10 putative *Pf*SET genes were found to be widely distributed over chromosomes 4-6, 8, 9, 11-13 as shown in Figure 1. A maximum of 2 genes were located on chromosome 12 (PF3D7_1214200 & PF3D7_1221000) and 13 (PF3D7_1322100 & PF3D7_1355300). Out of 10 *Pf*SET putative genes, only 5 genes were found to contain introns. Number of introns varied between a range of minimum 1 (PF3D7_1322100 & PF3D7_0910000) to maximum 6 (PF3D7_0508100) in these genes. Length of *Pf*SET proteins varied from a minimum of 178 amino acids (PF3D7_1214200) to a maximum of 6753 amino acids (PF3D7_0629700) as shown in Figure 2. There were large variations in isoelectric point (pI) values; ranging from acidic 4.2 (PF3D7_1115200) to basic 9.51 (PF3D7_1214200) and molecular weight from 20.7 kDa (PF3D7_1214200) to 796.1 kDa (PF3D7_0629700).

Table 1.Description of *Pf*SET genes

S. No.	Gene ID	Product description	Chromosome No. [Gene location]	Nucleotide length (bp)	Protein length (aa)	Molecular weight (Da)	No. of SET (SMART & pfam)	Isoelectric point (pI)	No. of Introns
1.	PF3D7_0629700	SET domain protein, putative (SET 1)	6 [Pf3D7_06_v3: 1,221,941 - 1,242,922 (+)]	20259	6753	796053	1	8.77	0
2.	PF3D7_1322100	SET domain protein, putative (SET 2)	13 [Pf3D7_13_ v3: 920,971 - 929,449 (+)]	7644	2548	300717	1	7.43	1
3.	PF3D7_0827800	SET domain protein, putative (SET 3)	8 [Pf3D7_08_v3: 1,198,858 - 1,206,057 (+)]	7197	2399	283552	1	8.3	5
4.	PF3D7_0910000	SET domain protein, putative (SET 4)	9 [Pf3D7_09_v3: 452,399 - 455,743 (-)]	3342	1114	131892	1	8.6	1
5.	PF3D7_1214200	histone-lysine N-methyltransferase, putative (SET 5)	12 [Pf3D7_12_ v3: 607,379 - 608,006 (-)]	534	178	20663	1	9.51	4

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

6.	PF3D7_1355300	histone-lysine N-methyltransferase, putative (SET 6)	13 [Pf3D7_13_ v3: 2,195,449 - 2,196,978 (+)]	1527	509	60024	1	8.82	0
7.	PF3D7_1115200	histone-lysine N-methyltransferase (SET 7)	11 [Pf3D7_11_ v3: 576,773 - 580,276 (+)]	2379	793	94294	1	4.2	0
8.	PF3D7_0403900	SET domain protein, putative (SET 8)	4 [Pf3D7_04_v3: 219,087 - 222,838 (+)]	3558	1186	142740	1	8.56	0
9.	PF3D7_0508100	SET domain protein, putative (SET 9)	5 [Pf3D7_05_v3: 331,680 - 336,704 (+)]	5022	1674	195630	0	8.91	6
10.	PF3D7_1221000	histone-lysine N-methyltransferase, putative (SET 10)	12[Pf3D7_12_v3: 836,912 - 843,901 (-)	6987	2329	271072	0	6	0

Information regarding product description, gene location, nucleotide sequence length, no. of introns, isoelectric point, molecular weight and amino acid sequence length is extracted from Plasmo DB. Confirmation of SET domain in each gene is based on SMART and Pfam database.

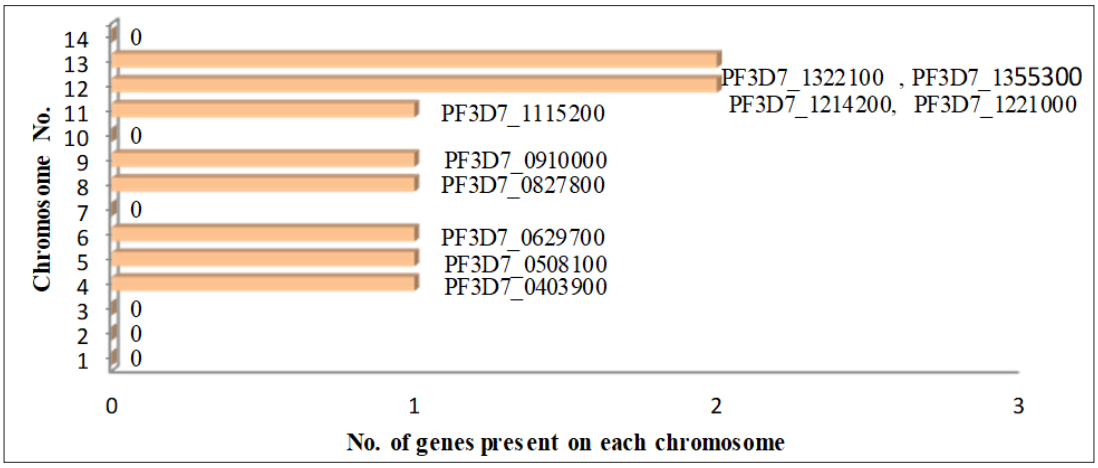


Figure 1.Distribution of *Pf*SET genes on the chromosomes of *P. falciparum*

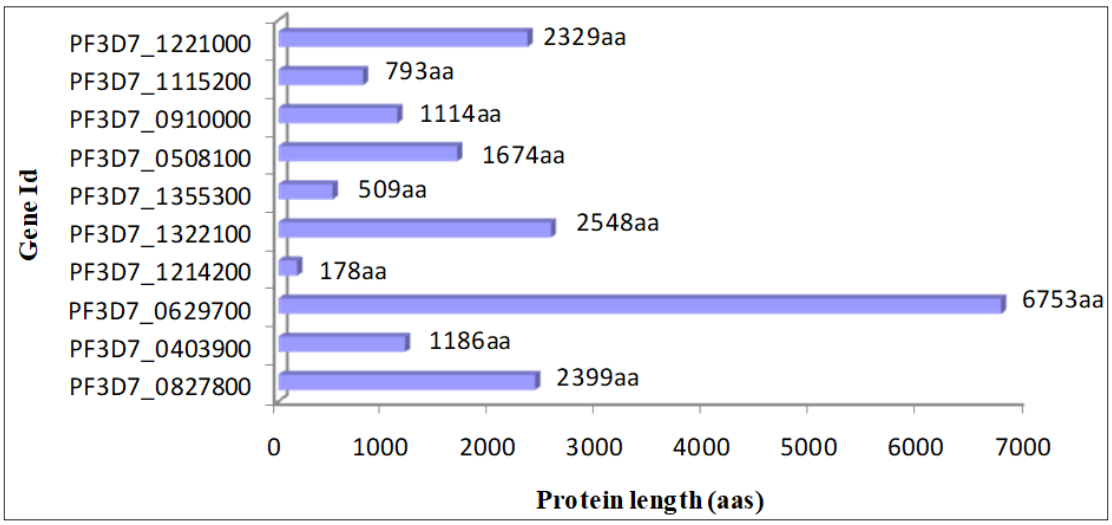


Figure 2.Graphical representation of protein sequence length of *Pf*SETs with their respective gene Ids at Y-axis

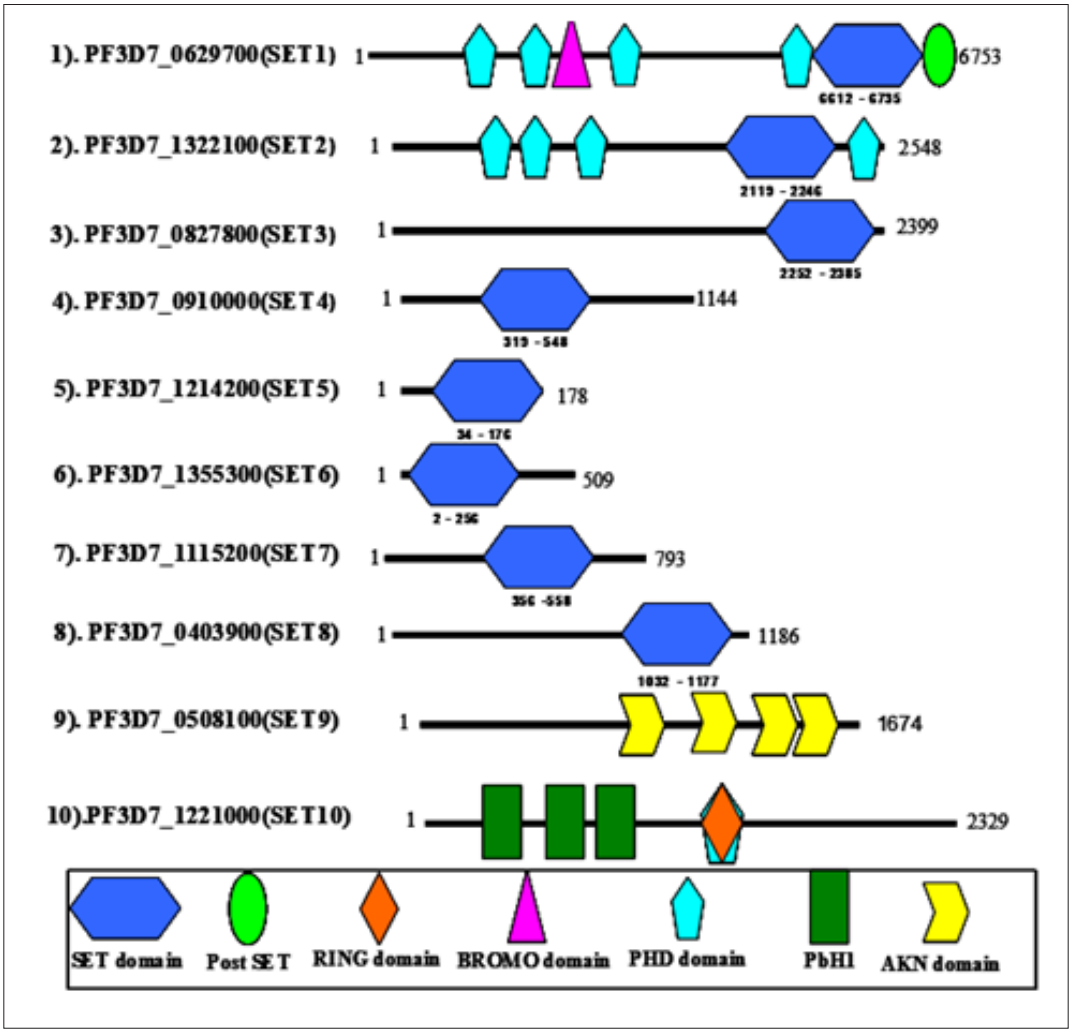
ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

Domain Architecture of *Pf*SET Proteins

The domain architecture of proteins encoded by *Pf*SET genes is based on the presence of SET domain and other additional domains as identified by SMART database. Out of total ten *Pf*SETs, six *Pf*SETs (PF3D7_0827800, PF3D7_0910000, PF3D7_1214200, PF3D7_1355300, PF3D7_1115200 and PF3D7_0403900) were found to encode only SET domain. Remaining four *Pf*SETs were found to have other domains besides SET domain as shown in Figure 3. Two *Pf*SETs (PF3D7_0629700, PF3D7_132200) were found to have 4 PHD domains additional to SET domain. PF3D7_0629700 was identified to encode one Bromo domain and one post SET domain besides PHD and SET domains. PF3D7_0508100 and PF3D7_1221000 were found to have no SET domain. But these were annotated as SET genes at PlasmoDB. *Pf*SET9 contains Ankyrin repeats (AKN domain) whereas *Pf*SET10 (PF3D7_1221000) harbors one PHD, one RING and three PbH1 domains.

Prediction of Subcellular Localization of *Pf*SET Proteins

*Pf*SET sequences were analyzed for putative signal sequences by MITOPROT, Euk-mpLoc 2.0 server and PATS and depicted in Figure 4. Out of eight *Pf*SETs bearing nuclear localization signals, six (55%) were found to be exclusive to this organelle as shown in Figure 4 (a). *Pf*SET9 (PF3D7_0508100) was found to be present in nucleus and cytoplasm both, whereas *Pf*SET5 (PF3D7_1214200) was found to be targeted to nucleus and mitochondria. *Pf*SET6 (PF3D7_1355300) was found to be cytoplasmic exclusively. No *Pf*SET was found to be exclusively mitochondrial or apicoplast targeted. However, *Pf*SET7 (PF3D7_1115200) was predicted to be shuttling between cytoplasm and mitochondria. Cellular localization of *Pf*SET proteins is represented in Figure 4 (b). The predicted nuclear localization for most of *Pf*SETs confirms their role in chromatin related processes.



The domain architecture is based on the presence of SET and other additional domains as identified by SMART and Pfam. Protein sizes are not to scale

Figure 3.Domain architecture *Pf*SET proteins

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

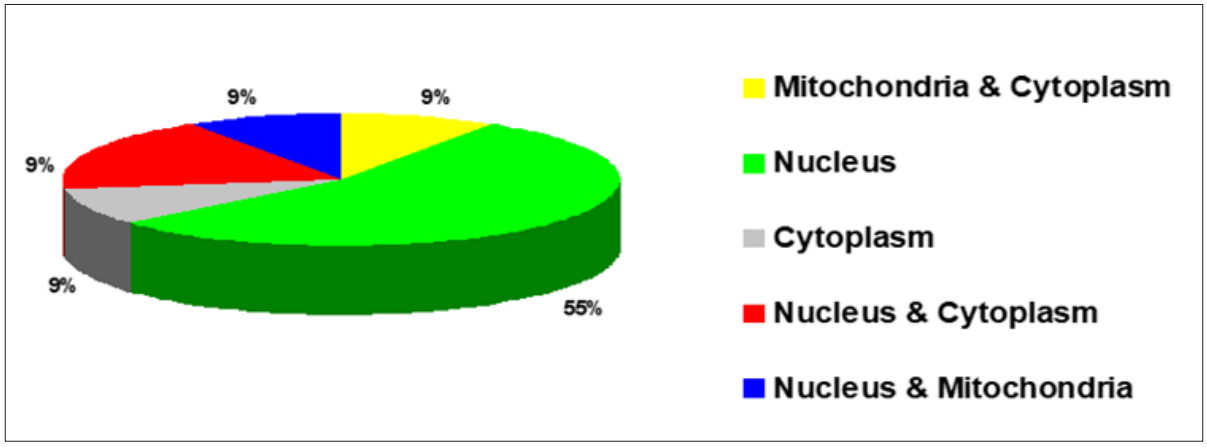


Figure 4(a).Percentage predicted distribution of *Pf*SETs in different organelles of the parasite

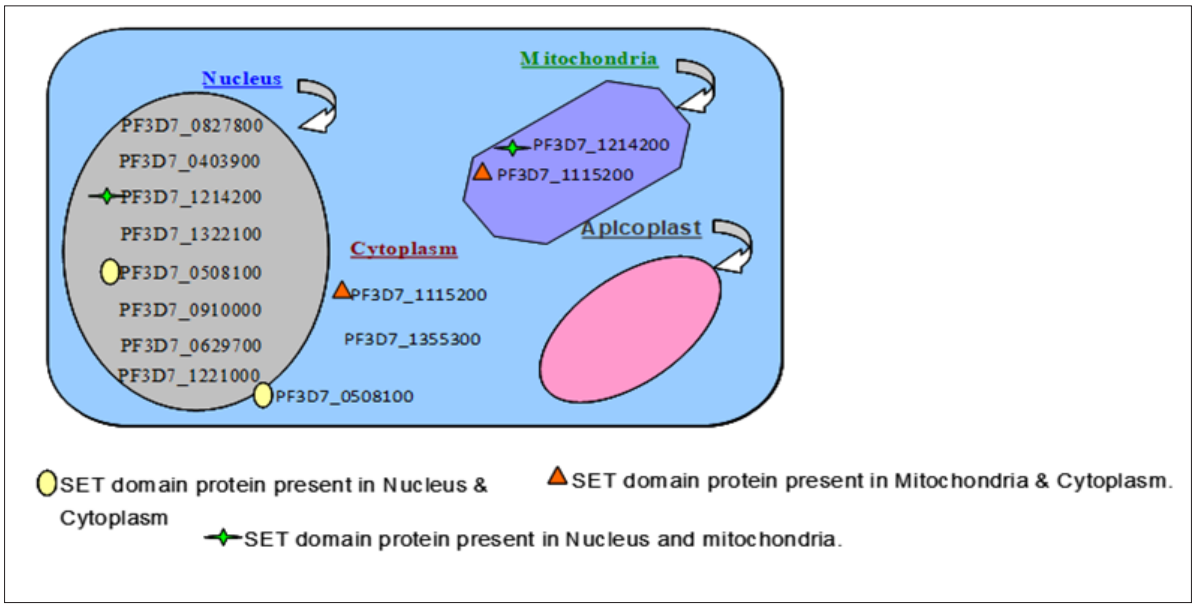


Figure 4(b).A schematic predicted cellular localization of *Pf*SET proteins

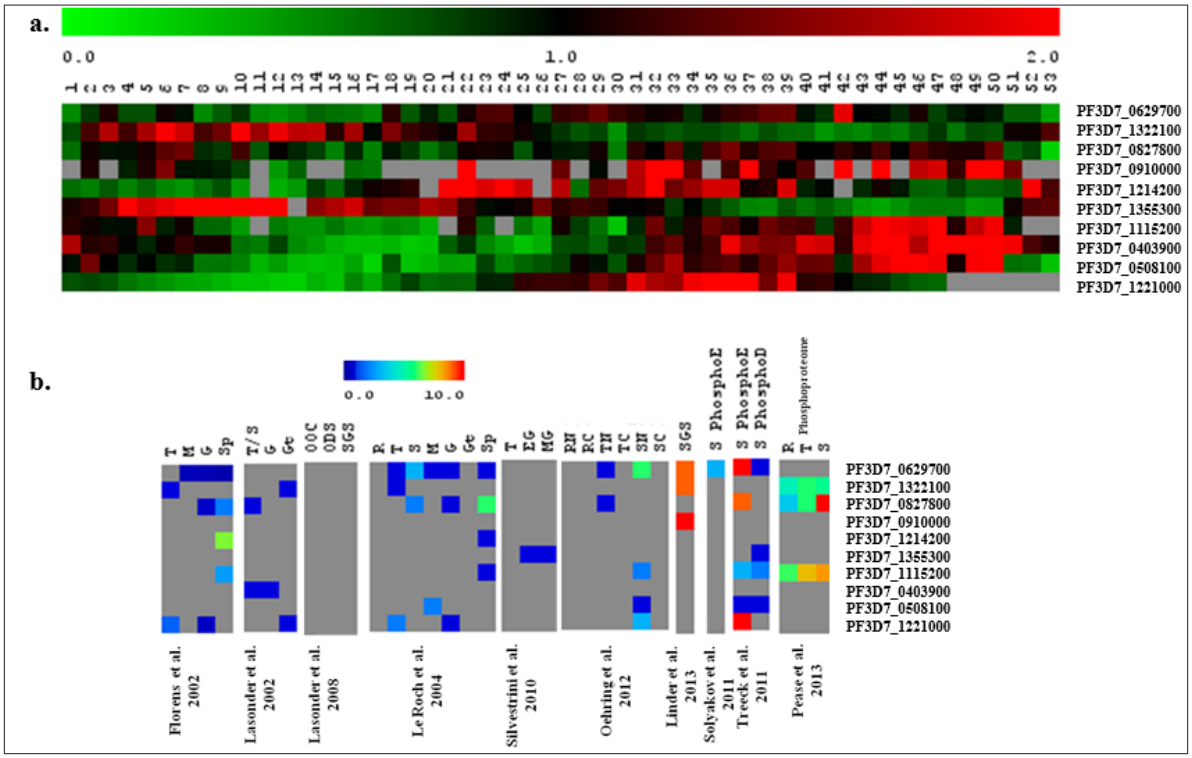
Expression Profiling of *Pf*SET Genes

In order to study expression of *Pf*SET genes during intraerythrocytic developmental cycle (IDC) of the malaria parasite life cycle, we analyzed transcriptome data by DeRisi group (Figure 5a).[31] DeRisi group represented the transcriptome of IDC for *P. falciparum* 3D7 at one hour time interval with 53 time points. The Proteome data of various studies[2,3,5,25-30] was compiled in Figure 5b.

PF3D7_0629700 was found to have constant expression profile during IDC according to transcriptome data. The protein of PF3D7_0629700 was expressed in T, S, M G and Sp. PF3D7_1322100 was identified to be upregulated in ring and having a constant expression profile in trophozoite and schizont stages at mRNA level and its protein was detected in R,T,S, Gt and SGS. PF3D7_0827800 gene was predicted to have constant expression throughout all three erythrocytic stages at transcript level. However, a real-time quantitative expression analysis of *Pf*SETs highlighted discrepancies in transcript abundance of PF3D7_0827800 as compared to microarray datasets[4,5,31] with a peak expression in schizont stage. The Proteomic datasets suggested its protein existence at R,T,S,G and Sp. PF3D7_0910000 protein was found to be SGS specific. PF3D7_1355300 was found to be upregulated in R at mRNA whereas its protein was restricted to S, EG, LG and Sp stages. Transcripts of PF3D7_1115200, PF3D7_0403900 and PF3D7_0508100 were found to be upregulated in S. PF3D7_1115200 protein expressed in R, T, S, and Sp. PF3D7_0403900 protein was detected at T/S and G only by Lasonder et al., 2002.[3] PF3D7_0508100 protein was found to be specific to S and M.

Expression profiling at transcript and protein level of *Pf*SET genes suggested a mixture of relationships between transcriptome and proteome datasets. Further, present analysis also revealed disparities between different proteome datasets.

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934



a) Heat map of Derisi microarray data of Llinas et al.[31] b) Heat map of proteome and phosphoproteome data retrieved from Florens et al.[2], Lasonder et al.[3], Le Roch et al.[5], Silvestrini et al.[25], Solyakov et al.[26], Treeck et al.[27], Oehring et al.[28], Lindner et al.[29], Pease et al.[30]. Different stages of P. falciparum are represented as: R, rings; T, trophozoites; S, schizonts; M, merozoites; G, gametocytes; Sp, sporozoites; Gt, gamete; EG, early gametocyte; MG, mature gametocyte; OOC, oocyst; ODS, oocyst derived sporozoites; SGS, salivary gland sporozoites; phospho E, phospho-enriched; phosphoD, phospho-depleted; N, nuclear; and C, cytoplasm.

Figure 5.Expression Profiling of *Pf*SETs

Table 2.List of human homologs of *Pf*SET proteins

S. No.	*P. falciparum* Gene ID	Product Description	Protein sequence length (aa)	*Homo sapiens* Gene ID	Product description	Protein sequence length (aa)
1.	PF3D7_0629700	SET domain protein, putative (SET 1)	6753	NP_005924.2	Histone-lysine N-methyltransferase MLL1	3969
2.	PF3D7_1322100	SET domain protein, putative (SET 2)	2548	NP_075447.1	Histone-lysine N-methyltransferase NSD3	1437
3.	PF3D7_1355300	histone-lysine N-methyltransferase, putative (SET 6)	509	NP_073580.1	SET and MYND domain-containing protein 3	428
4.	PF3D7_1115200	SET domain protein, putative (SET 7)	793	NP_006053.1	SET and MYND domain-containing protein 5	418
5.	PF3D7_0403900	SET domain protein, putative (SET 8)	1186	NP_065115.3	N-lysine methyltransferase SETD8	393

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

Prediction of Human Homologs of *Pf*SETs

Prediction of homologs of *Pf*SET proteins in *Homo sapiens* was carried out using BLASTp of *Pf*SET protein sequences against HPRD (Human Protein Reference Database). Out of total 10 *Pf*SET genes, we could predict human homologs for 5 *Pf*SET genes (Table 2). All *Pf*SETs are longer in length than their human homologs highlighting parasite specific insertions. Further, it will be interesting to compare domain structure and sequences of these *Pf*SET proteins with their human homologs to identify parasite specific features.

Multiple Sequence Alignment of *Pf*SET Genes

SET domain harbors four signature motifs- motif 1 (G-X-G), motif II (YXG), motif III (RFINHXCXPN) and motif IV (ELXFDY).[34-37] By multiple sequence alignment, we were able to recognize all four motifs I, II, III and IV of SET domain as shown in Figure 6. Motif 1 and II were found to be less conserved as compared to motifs III and IV. Motifs III and IV form a pseudo knot fold which brings the two most-conserved sequence motifs III (RFINHXCXPN) and IV (ELXFDY) of the SET domain together to form an active site.[35] The conserved residues (G-X-G) of motif 1 and NH residues of the motif III are involved in hydrogen bonding and van der Waals interactions with the cofactor AdoMet.[36] Further, Y residue of motif IV was found to be conserved in the lysine binding cleft which form hydrogen-bond with amino group of K residue and align it for a methyltransfer with AdoMet.[37]

Figure 6.Multiple sequence alignment of *Pf*SET domain sequences highlighting conserved features

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

Generation of Sequence Logo of *Pf*SET

A sequence logo for SET domain of *P. falciparum* was generated by Web logo based on multiple sequence alignment of all 8 SET domain sequences identified in *P. falciparum*. Generated *Pf*SET domain sequence logo is shown in Figure 7. The overall height of each stack is proportional to the residue conservation at that position and the height of each letter corresponds to frequency of particular residue at that position. The logo shows the level of conservation of four signature motif consensus sequences as predicted by multiple sequence alignment. In addition to these motifs conservation of some more residues like K (lysine), Y (tyrosine), A (alanine), R (arginine) and D (aspartic acid) is also revealed.

Phylogenetic Analysis of *Pf*SETs

An un-rooted neighbor-joining Phylogenetic tree for 10 *Pf*SET protein sequences was constructed using Phylip 3.69 and was visualized with MEGA 5.05 program (Figure 8). Phylogenetic analysis revealed that PfSET1, PfSET9, PfSET4, PfSET5 and PfSET10 are grouped together in one calde. PfSET7 and PfSET8 are closely related to each other and grouped together. However, PfSET3 and PfSET6 revealed distant relationships.

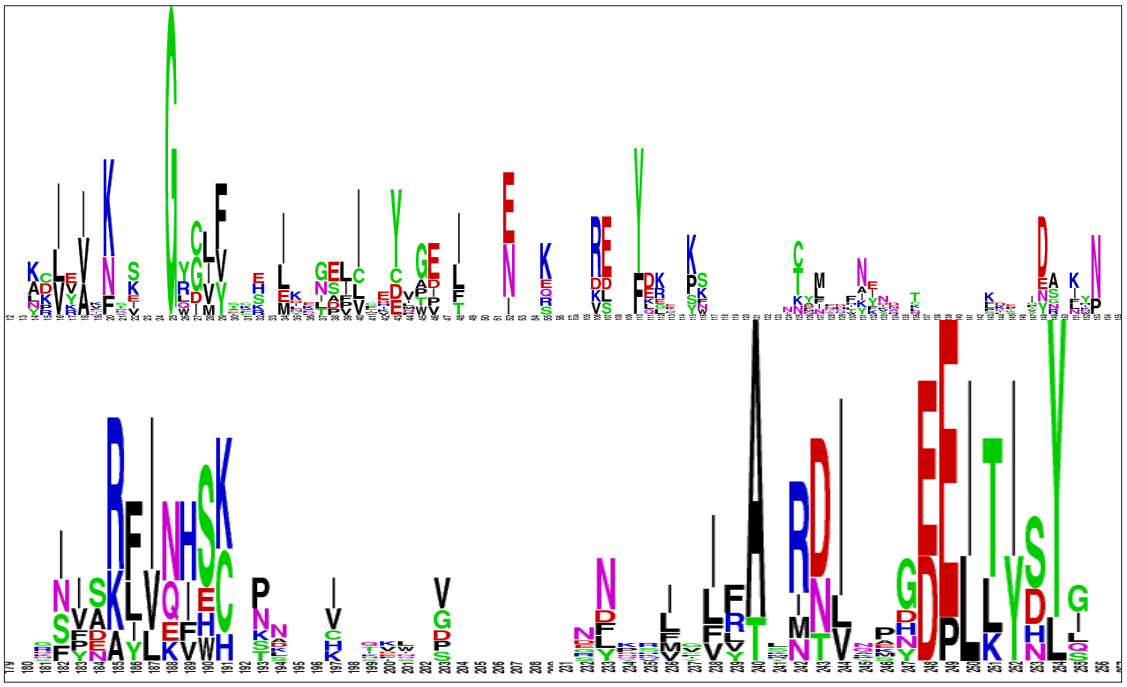


Figure 7.HMM logo of *Pf*SET domain revealing conserved residues. The logo was generated using Web logo

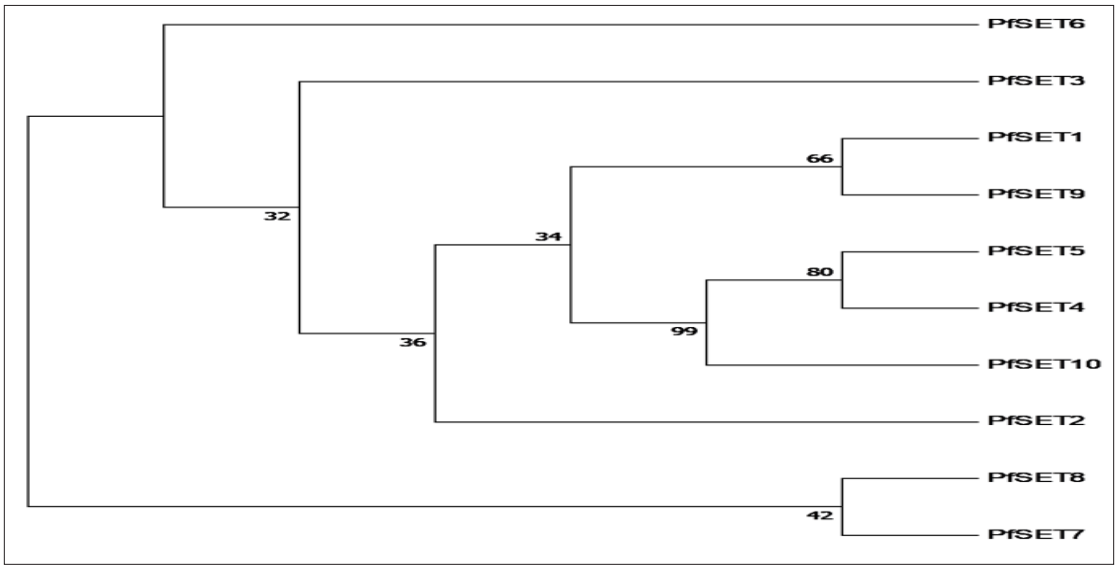


Figure 8.Unrooted phylogenetic tree of *Pf*SETs

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

Conclusion

The present study provides detailed bioinformatics analysis of *Pf*SET family. In this study, we identified 10 *Pf*SET genes in *P. falciparum*. Further, we carried out detailed analysis of their domain composition, physiochemical properties, subcellular localization, expression profiling and phylogenetic relationships. The work highlighted the unique domain composition and conservation of important catalytic residues in *Pf*SETs. Transcriptome and proteome profiling of *Pf*SETs exhibited a mixture of linear and nonlinear relationships and revealed stage specific *Pf*SETs. In essence, genome-wide analysis of *Pf*SET family in *P. falciparum* will provide a platform for further experimental studies.

Acknowledgement

MK, PC thankfully acknowledge the Maharshi Dayanand University, Rohtak, Haryana for University Research Fellowship. RG acknowledges the UGC and DST funding.

Conflict of Interest: None

References

1. World Malaria Report, 2018. Available from: https://www.who.int/malaria/publications/world-malaria-report-2018/report/en/.
2. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD et al. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 2002; 419(6906): 520-526. Available from: https://www.nature.com/articles/nature01107 [PubMed/ Google Scholar].
3. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW et al. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 2002; 419(6906): 537-542. Available from: https://www.nature.com/articles/nature01111 [PubMed/ Google Scholar].
4. Bozdech Z, Llinás M, Pulliam BL, Wong ED, Zhu J, DeRisi JL et al. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 2003; 1(1): E5. Available from: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0000005 [Google Scholar].
5. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 2003; 301(5639): 1503-1508. Available from: https://science.sciencemag.org/content/301/5639/1503.long [PubMed/ Google Scholar].
6. Fischle W, Wang Y, Allis CD. Histone and chromatin cross-talk. *Curr Opin Cell Biol* 2003; 15: 172-183. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0955067403000139?via%3Dihub [PubMed/ Google Scholar].
7. Cui L, Miao J. Chromatin-Mediated Epigenetic Regulation in the Malaria Parasite Plasmodium falciparum. *Eukaryot Cell* 2010; 9(8): 1138-1149. [PubMed/ Google Scholar].
8. Murray K. The occurrence of N-methyl lysine in histones. *Biochemistry* 1964; 3(1): 10-15. Available from: https://pubs.acs.org/doi/abs/10.1021/bi00889a003 [Google Scholar].
9. Schneider R, Bannister A, Kouzarides T. Unsafe SETs: Histone methyltransferases and cancer. *Trends Biochem Biol* 2002; 27(8): 396-402. [PubMed/ Google Scholar].
10. Kouzarides T. Chromatin modifications and their function. *Cell* 2007; 128: 693-705. [PubMed/ Google Scholar].
11. Jenuwein T, Laible G, Dorn R, Reuter G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci* 1998; 54(1): 80-93. Available from: https://link.springer.com/article/10.1007%2Fs000180050127 [PubMed/ Google Scholar].
12. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000; 406: 593-599. Available from: https://www.nature.com/articles/35020506 [PubMed/ Google Scholar].
13. Lei L, Zhou SL, Ma H, Zhang LS. Expansion and diversification of the SET domain gene family following whole-genome duplications in Populus trichocarpa. *Evolutionary Biology* 2012; 12(51): 1471-2148. Available from: https://link.springer.com/article/10.1186/1471-2148-12-51 [PubMed/ Google Scholar].
14. Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G. The protein encoded by the *Drosophila* position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J* 1994; 13(16): 3822-3831. [PubMed/ Google Scholar].
15. Jones RS, Gelbart WM. The *Drosophila* Polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to trithorax. *Mol Cell Biol* 1993; 13(10): 6357-6366. Available from: https://mcb.asm.org/content/13/10/6357.long [PubMed/ Google Scholar].
16. Stassen MJ, Bailey D, Nelson S, Chinwalla V, Harte PJ. The *Drosophila* trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mech Dev* 1995; 52(2-3): 209-223. Available from: https://www.sciencedirect.com/science/article/pii/092547739500402M?via%3Dihub [PubMed/ Google Scholar].

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

17. Coetzee N, Sidoli S, van Biljon R, Painter H, Llinás M, Garcia BA et al. Quantitative chromatin proteomics reveals a dynamic histone post-translational modification landscape that defines asexual and sexual *Plasmodium falciparum* parasites. *Sci Rep* 2017; 7(1): 607. Available from: https://www.nature.com/articles/s41598-017-00687-7 [PubMed/ Google Scholar].
18. Read DF, Cook K, Lu YY, Le Roch KG, Noble WS. Predicting gene expression in the human malaria parasite *Plasmodium falciparum* using histone modification, nucleosome positioning, and 3D localization features. *PLoS Comput Biol* 2019; 15(9): e1007329. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007329 [PubMed/ Google Scholar].
19. Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K et al. PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature* 2013; 499(7457): 223-227. Available from: https://www.nature.com/articles/nature12361 [PubMed/ Google Scholar].
20. Aurrecoechea C, Brestelli J, Brunk BP, et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* 2009; 37: D539-543. Available from: https://academic.oup.com/nar/article/37/suppl_1/D539/1012097 [PubMed/ Google Scholar].
21. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* 2018; 46: D493-D496. Available from: https://academic.oup.com/nar/article/46/D1/D493/4429069 [PubMed/ Google Scholar].
22. Claros MG, Vincens P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* 1996; 241(3): 779-786. Available from: https://febs.onlinelibrary.wiley.com/doi/full/10.1111/j.1432-1033.1996.00779.x?sid=nlm%3Apubmed [PubMed/ Google Scholar].
23. Chou KC, Shen HB. A new method for predicting the sub-cellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. *PLoS ONE* 2010; 5(4): e9931. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009931 [PubMed/ Google Scholar].
24. Zuegge J, Ralph S, Schmuker M, McFadden GI, Schneider G. Deciphering apicoplast targeting signals-feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins. *Gene* 2001; 280(1-2): 19-26. [PubMed/ Google Scholar].
25. Silvestrini F, Lasonder E, Olivieri A, Camarda G, van Schaijk B, Sanchez M et al. Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 2010; 9(7): 1437-48. Available from: https://www.mcponline.org/content/9/7/1437.long [PubMed/ Google Scholar].
26. Solyakov L, Halbert J, Alam MM, Semblat JP, Dorin-Semblat D, Reininger L et al. Global kinomic and phospho-proteomic analyses of the human malaria parasite *Plasmodium falciparum*. *Nat Commun* 2011; 2: 565. Available from: https://www.nature.com/articles/ncomms1558 [PubMed/ Google Scholar].
27. Treeck M, Sanders JL, Elias JE, Boothroyd JC. The phosphoproteomes of *Plasmodium falciparum* and *Toxoplasma gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* 2011; 10(4): 410-419. [PubMed/ Google Scholar].
28. Oehring SC, Woodcroft BJ, Moes S, Wetzel J, Dietz O, Pulfer A et al. Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum*. *Genome Biol* 2012; 13(11): R108. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-11-r108 [PubMed/ Google Scholar].
29. Lindner SE, Swearingen KE, Harupa A, Vaughan AM, Sinnis P, Moritz RL et al. Total and putative surface proteomics of malaria parasite salivary gland sporozoites. *Mol Cell Proteomics* 2013; 12(5): 1127-1143. Available from: https://www.mcponline.org/content/12/5/1127.long [PubMed/ Google Scholar].
30. Pease BN, Huttlin EL, Jedrychowski MP, Talevich E, Harmon J, Dillman T et al. Global analysis of protein expression and phosphorylation of three stages of *Plasmodium falciparum* intraerythrocytic development. *J Proteome Res* 2013; 12(9): 4028-4045. Available from: https://pubs.acs.org/doi/10.1021/pr400394g [PubMed/ Google Scholar].
31. Llinás M, Bozdech Z, Wong ED, Adai AT, DeRisi JL. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* 2006; 34(4): 1166-1173. Available from: https://academic.oup.com/nar/article/34/4/1166/1337467 [PubMed/ Google Scholar].
32. Prasad TSK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S et al. Human Protein Reference Database - 2009 Update. *Nucleic Acids Res* 2009; 37: D767-D772. Available from: https://academic.oup.com/nar/article/37/suppl_1/D767/1019294 [PubMed/ Google Scholar].
33. Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.69. Department of Genome Sciences, University of Washington, Seattle 2009.
34. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K et al. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* 2002; 12(12): 1052-1058. [PubMed/ Google Scholar].

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934

35. Cui L, Fan Q, Cui L, Miao J. Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*. *International Journal for Parasitology* 2008; 38(2008): 1083-1097. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0020751908000210 [PubMed/ Google Scholar].
36. Dillon SC, Zhang X, Trievel RC, et al. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology* 2005; 6: 227. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2005-6-8-227 [Google Scholar].
37. Cheng X, Collins RE, Zhang X. Structural and sequence motifs of protein (histone) methylation enzymes. *Annu Rev Biophys Biomol Struct* 2005; 34: 267-294. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev.biophys.34.040204.144452?rfr_dat=cr_pub%3Dpubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&journalCode=biophys.3 [PubMed/ Google Scholar].

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.201934