Serdica J. Computing **12** (2018), No 3, 131–162

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

HUMAN AND RODENT MALARIA PARASITES DIFFER IN THEIR HISTONE METHYLATION MACHINERY— A BIOINFORMATIC STUDY

Malgorzata Perycz, Piotr Zielenkiwiecz, Szymon Kaczanowski

ABSTRACT. Gene expression in *Plasmodium* can be regulated epigenetically similarly to the other eukaryotic species. Methylation is one of the histone modifications that affect chromatin structure and influence gene expression. Recently it was reported that rodent *Plasmodium* species lack orthologs to [PlasmoDB:PF3D7_0809900], a putative histone demethylase. We take advantage of progress in genomics and perform a whole genome analysis.

Based on sequence and phylogenetic analysis of putative enzymes involved in histone methylation in *Plasmodium*, we show that rodent malaria parasites lost not only a putative histone demethylase but also a putative histone methylase. *In silico* prediction showed that these enzymes were likely to target the same lysine in histone H3, and suggested the whole histone-modifying module was lost in rodent *Plasmodium* species. The evolutionarily old histone modification machinery differs in rodent and primate *Plasmodium* species, suggesting that change of the histone code during malaria evolution took place early in *Plasmodium* speciation.

ACM Computing Classification System (1998): J.3.

Key words: Histone code, Plasmodium, malaria, SET domain, JmjC, Jumonji, histone demethylase, histone methylase, evolution.

1. Background. Malaria is a still neglected but widespread and deadly infectious disease. Annually, over 250 million people become infected and over 800 000 of them die [1]. The causative agent of malaria is a protozoan of the genus *Plasmodium* [reviewed in 2]. The rodent malaria parasites *P. berghei*, *P. yoelii*, and *P. chabaudi* are natural parasites of thicket rats in central Africa and are widely used in malaria research.

The life cycle of *Plasmodium* is complex and features multiple developmental forms in vertebrate and invertebrate hosts. The microarray studies reveal substantial variation in gene expression that closely correlates with the diversity of phenotypes [3, 4] and may be subject to control by epigenetic elements. The virulence of *Plasmodium* is dependent on a set of gene families that control antigenic variation and have been linked to host immune evasion [reviewed in 5]. Recent studies have implicated that chromatin conformation and histone modification influence the expression of one such family in *P. falciparum*—the *var* genes [6-11] and hence contribute to regulation of a parasite's virulence.

Plasmodium displays a paucity of protein domains associated with specific transcription factors and contains mostly basal chromatin modifying factors that are generally conserved among eukaryotes [12]. Based on the microarray data for the *P. falciparum* asexual stages, Bozdech et al. [3] suggested that a small number of transcription factors with overlapping binding site specificities could account for the mechanical character of transcriptional control in the parasite. Hence, the regulation of gene expression on the chromatin conformation level may play an especially significant role in achieving specific transcription in *Plasmodium*. The histone code hypothesis presumes that gene expression is critically influenced by chemical modifications to histone proteins. These covalent histone modifications include methylation, acetylation, phosphorylation, sumoylation and ubiquitination, and have been shown to alter chromatin structure and gene expression [13]. Histone methylation occurs mainly on lysine and arginine residues of histone N-terminal tails but can also occur within the globular core of the histone with the action of Dot1 enzyme [14-17]. Histone lysines can be mono-, di-, and trimethylated [13], and histone arginines can be mono- and di-methylated [18]. Each distinct methyl state may have different biological functions. For

example, a dimethylation of lysine residue 9 of histone H3 (H3K9me2) and trimethylation of lysine residue 27 of histone H3 (H3K27me3) negatively correlate with gene activity, whereas di- and trimethylation of histone 3 lysine residue 4 (H3K4me3) and histone 3 lysine residue 36 (H3K36me3) are associated with gene expression [19, 20].

Histone methylation is accomplished through the action of histone methyltransferases (HMTs), a category of proteins which have been described and investigated in many systems [21, 22].

The known histone methylases can be clustered into three groups: SET-domain methyltransferases acting towards lysines within the histone tails, non-SET domain methyltransferases featuring single enzyme Dot1 homologs, and arginine-specific methyltransferases. Histone methyltransferases containing a SET domain (Suvar3-9, Enhancer-of-zeste, Trithorax domain) compose the biggest family of histone methylases. SET is an active domain of approx. 130 amino acids, known mostly for its interaction with the histone lysines. Proteins containing SET domains are involved in both activating [23] and silencing [24] of gene expression [for review see: 20, 25, 26]. Within proteins, SET domains are often accompanied by other identifiable domains, including domains interacting with nucleosomes and DNA. Histone methylation was thought to be an irreversible modification until the discovery of histone demethylases [27]. The first discovered group of histone demethylases were amine oxidases. In humans this family consists of two known members: Lysine-specific histone demethylase 1A (product of KDM1A gene, yeast homolog: LSD1) and Lysine-specific histone demethylase 1B (KDM1B, or Lysine-specific histone demethylase 2). Lysine-specific histone demethylase 1A was shown to remove mono- and dimethyl groups from histone 3 lysine residue 4 (H3K4) [27], and methyl groups from histone 3 lysine residue 9 (H3K9) [28], thereby acting as a coactivator or a corepressor, depending on the context. Murine lysine-specific histone demethylase 1B was shown to demethylate 'Lys-4' of histone H3, thereby acting as a corepressor, and to be required for *de novo* DNA methylation of a subset of imprinted genes during oogenesis [29, 30]. The second family of histone demethylases is constituted by Jumonji proteins [31, 32]. It has about 30 members described in humans. They contain conserved JmjC domains, which are crucial for their demethylase acitivity [33, 34]. The

various JmjC domain containing proteins target specific histone lysines at different methylation states [19]. For example, human lysine-specific demethylase 2A specifically demethylates 'Lys-36' of histone H3, preferentially dimethylated residue while it has weak or no activity for mono- and trimethylated H3 'Lys-36 [33]. Another Lys-4 H3 demethylase is yeast histone demethylase JHD2, which targets trimethylated 'Lys-4' of histone H3 [35]. A bioinformatic comparative proteome study in rodent and primate malaria parasites [36] led to identification of a deletion of orthologs of putative histone demethylase H3K36 [PlasmoDB:PF3D7_0809900)] in rodent *Plasmodium* species. On the basis of previous observation that protein modules seem to be inactivated commonly in evolution [37, 38], we asked if any methylases were also lost in rodent malaria parasites and searched for possible candidates in rodent and primate malaria parasites proteomes.

In this paper we present results of our analysis, showing that rodent species of *Plasmodium* lost not only the aforementioned H3K36 demethylase, but also a SET-domain containing putative histone methylase. This whole histone modification module missing may indicate a change in chromatin regulation between primate and rodent *Plasmodium* species, and increases our understanding of the parasites histone code.

2. Results. The class of putative histone demethylases was lost in rodent *Plasmodium* species early in evolution.

Frech and Chen [36] in their global study identified a putative histone H3-lysine-36 demethylase as present in primate but lost in rodent *Plasmodia* (*P. falciparum* [PlasmoDB:PF3D7_0809900]). We constructed phylogenetic trees for both amino oxidase domain containing demethylases (Fig. 1) and Jumonji domain containing demethylases (Fig. 2) in *Plasmodium*. The protein sequences used for this analysis are listed in Tables 1 and 2 and Supplementary Table S1 and S2. Out of the identified plasmodial orthologs of known histone demethylases, only one type of demethylase was present in human *Plasmodium* while it is nonexistent in rodent *Plasmodium* (Fig. 2, Table 2). The tree suggests that this protein was lost early in the evolution of *Plasmodium*.

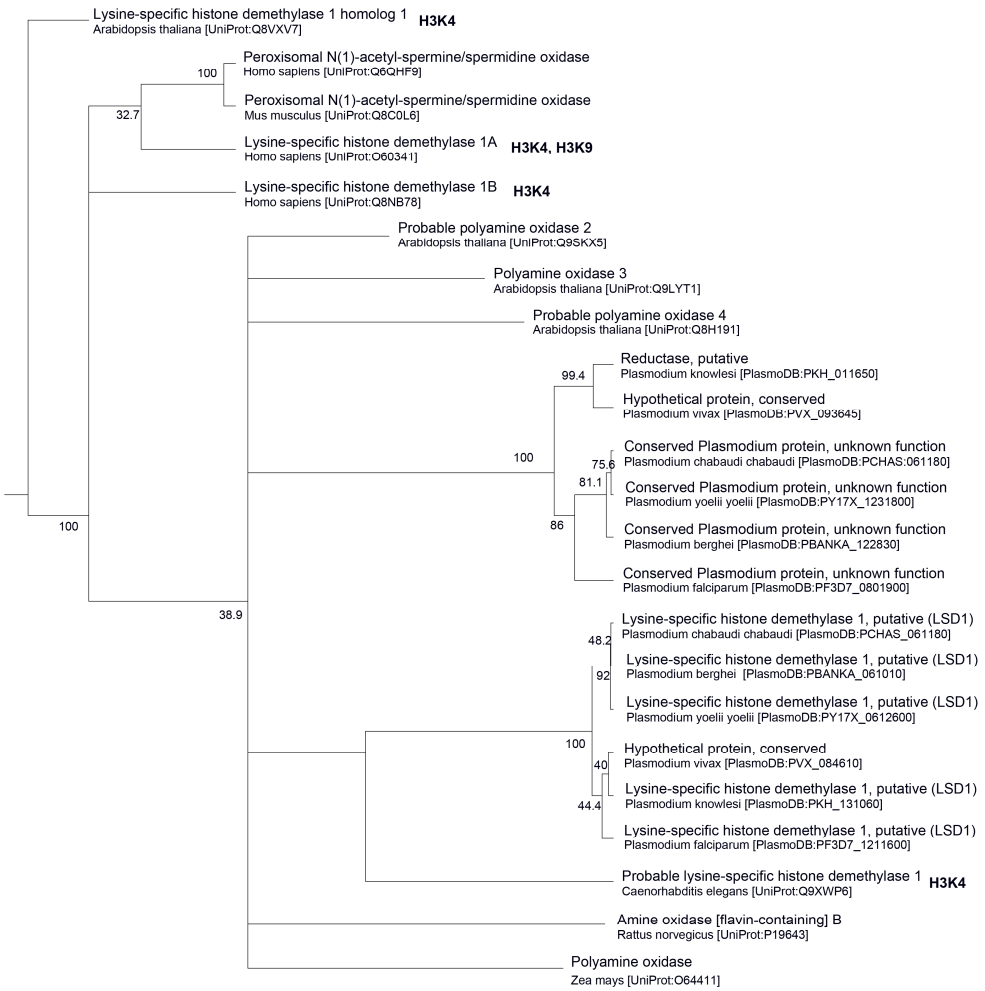

Fig. 1. Phylogenetic analysis of aminooxidase domain-containing proteins in *Plasmodium* and their homologs in other organisms.

Protein sequences were derived from NCBI, Uniprot and PlasmoDB databanks. Demethylation activity indicated where known. Aminooxidase domains were extracted from sequences using Pfam and aligned using Muscle algorithm. Alignments were edited in Jalview to remove gaps. The phylogenetic tree was calculated using protein parsimony algorithm (ProtPars, Phylip). The evolutionary distances were assessed using Evolutionary Clock (Phylip 3.67, Kitsch), based on a distance matrix computed from aligned protein sequences using Phylip 3.67: Protdist.

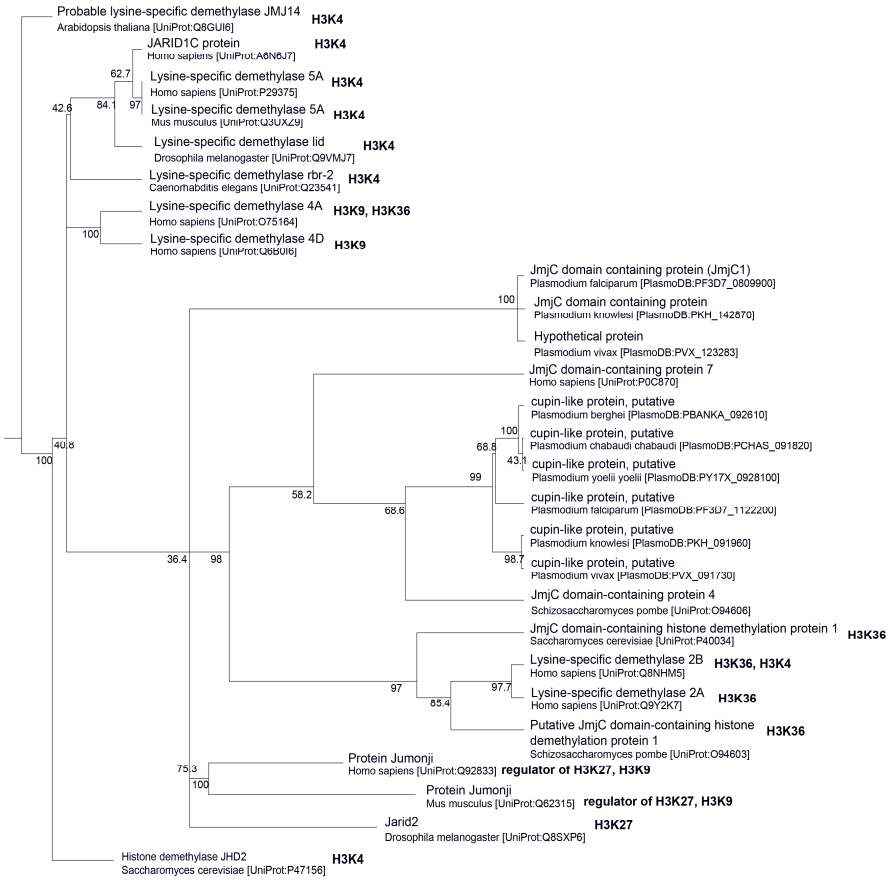

Fig. 2. Phylogenetic analysis of JmJC/cupin domain-containing proteins in Plasmodia and their homologs in other organisms.

Protein sequences were derived from NCBI, Uniprot and PlasmoDB databanks. Demethylation activity indicated where known. JmJC/cupin domains were extracted from sequences using Pfam and aligned using Muscle algorithm. Alignments were edited in Jalview. The phylogenetic tree was calculated using protein parsimony algorithm (ProtPars, Phylip). The evolutionary distances were assessed using Evolutionary Clock (Phylip 3.67, Kitsch), basing on a distance matrix computed from aligned protein sequences using Phylip 3.67: Protdist.

A class of SET-domain containing proteins that are putative histone methylases was lost in rodent Plasmodium species.

Table 1. Plasmodial homologs of human histone demethylases with an aminooxidase domain

Human histone demethylase	**Lysine-specific histone demethylase 1A**	**Lysine-specific histone demethylase 2A**
Gene names	KDM1A (AOF2/BHC110/LSD1)	KDM1B (AOF1/LSD2)
Specificity	H3K4me2/me1 H3K9me2/me1	H3K4me2/me1
P. falciparum homolog	PlasmoDB:PF3D7_0801900	PlasmoDB:PF3D7_1211600
P. vivax homolog	PlasmoDB:PVX_093645	PlasmoDB:PVX_084610
P. knowlesi homolog	PlasmoDB:PKH_011650	PlasmoDB:PKH_131060
P. berghei homolog	PlasmoDB:PBANKA_122830	PlasmoDB:PBANKA_061010]
P. chabaudi homolog	PlasmoDB:PCHAS_122900	PlasmoDB:PCHAS_061180]
P. yoelii homolog	PlasmoDB:PY17X_1231800	PlasmoDB:PY17X_0612600]
Putative histone demethylase domain	aminooxidase	amionooxidase

Table 2. Plasmodial homologs of human histone demethylases with JmjC domain

Human histone demethylase Gene names	JMJD7	KDM2A/ JHDM1A/ FBXL11	1) JHDM3A-D/JMJD2A-D 2) JARID1A/RBP2 JARID1B/PLU1 JARID1C/SMCX JARID1D/SMCY 3) JARID2
Histone demethylase activity	undetermined	H3K36me2/me1	1) H3K9me3/ me2H3K36me3/me2; 2) H3K4me3/me2 3) none (regulator of H3K9 and H3K27 methylation status)
P. falciparum homolog	PlasmoDB: PF3D7_1122200	*PlasmoDB: PF3D7_1123200	[PlasmoDB: PF3D7_0809900]
P. vivax homolog	*PlasmoDB: PVX_091730	*PlasmoDB: PVX_091775	PlasmoDB: PVX_123283
P. knowlesi homolog	PlasmoDB: PKH_091960	*PlasmoDB: PKH_092070	PlasmoDB: PKH_142870
P. berghei homolog	PlasmoDB: PBANKA_092610	PlasmoDB: PBANKA_092510	none
P. chabaudi homolog	PlasmoDB: PCHAS_091820	[PlasmoDB: PCHAS_091930]	none
P. yoelii homolog	*PlasmoDB: PY17X_0928100	[PlasmoDB: PY06315]	none
Putative histone demethylase *domain*	**cupin**	**none**	**JmjC**

** protein identified in orthology search in Plasmo DB using a Plasmodial homolog as a query. Plasmodial proteins found in a homology search with Lysine-specific demethylase 2A as a query did not contain a JmjC domain and were excluded from further phylogenetic analysis.*

Table S1. Aminooxidase domain containing proteins used in this study and their specificity towards methylated histones

Organism	**Accesion number (UniProt/NCBI/ PlasmoDB)**	**Protein name**	**Specificity/ Function**	**Citation**
H. sapiens	[UniProt:Q6QHF9]	Peroxisomal N(1)-acetyl-spermine/ spermidine oxidase	Flavoenzyme which catalyzes the oxidation of N(1)-acetylspermine to spermidine and is thus involved in the polyamine back-conversion.	[34]
H. sapiens	[UniProt:Q9NWM0]	Spermine oxidase/Poly amine oxidase 1 (PAO1)	Flavoenzyme which catalyzes the oxidation of spermine to spermidine. / Not determined	[35]
H. sapiens	[UniProt:Q8NB78-2]/ [UniProt:A2A2C6]	lysine-specific histone demethylase 1B (KDM1B) /oxidase (Flavin containing) domain 1	H3K4me2/me1	By similarity
H. sapiens	[UniProt:O60341]	Lysine-specific histone demethylase 1A	H3K4me2/me1 and H3K9	[25, 26, 36-38]
M. musculus	[UniProt:Q4GX45]	Paox	Peroxisomal N1-acetyl-spermine/spermidine oxidoreductase	[39]

M. musculus	[UniProt:Q8C0L6]	Peroxisomal N(1)-acetyl-spermine/ spermidine oxidase	Flavoenzyme which catalyzes the oxidation of N(1)-acetylspermine to spermidine. Can also oxidize N(1)-acetylspermidine to putrescine.	[40]
M. musculus	[UniProt:Q99K82]	Spermine oxidase	Flavoenzyme which catalyzes the oxidation of spermine to spermidine. Can also use N(1)-acetylspermine and spermidine as substrates	[41]
R. norvegicus	[UniProt:P19643]	Amine oxidase [flavin-containing] B	Catalyzes the oxidative deamination of biogenic and xenobiotic amines and has important functions in the metabolism of neuroactive and vasoactive amines in the central nervous system and peripheral tissues. MAOB preferentially degrades benzylamine and phenylethylamine.	[42]
S. pombe	[UniProt:Q9Y802]	Lysine-specific histone demethylase 1	H3K9me2 and H3K4me2	[43-45]
D. melano-gaster	[UniProt:P18487]	Protein anon-37Cs	Has a non-vital function; no data	[46]

D. melanogaster	[UniProt:Q9VW97]	Possible lysine-specific histone demethylase 1	demethylates H3K4me2 and H3K4me1 and facilitates subsequent H3K9 methylation by SU(VAR)3-9	[47, 48]
C. elegans	[UniProt:Q9XWP6]	Probable lysine-specific histone demethylase 1	Probable histone demethylase that specifically demethylates 'Lys-4' of histone H3 H3K4me2/1	By similarity; The C.elegans sequencing consortium, 1998
A. thaliana	[UniProt:Q9SKX5]	Probable polyamine oxidase 2	Flavoenzyme that catalyzes the oxidation of the secondary amino group of polyamines	By similarity
A. thaliana	[UniProt:Q9LYT1]	Polyamine oxidase 3	Flavoenzyme that catalyzes the sequential oxidation of spermine to spermidine, and of spermidine to putrescine.	[49]
A. thaliana	[UniProt:Q8H191]	Probable polyamine oxidase 4	Flavoenzyme that catalyzes the oxidation of the secondary amino group of spermine.	[50]
A. thaliana	[UniProt:Q8VXV7]	Lysine-specific histone demethylase 1 homolog 1	Probable histone demethylase that reduces the levels of histone H3 'Lys-4' methylation in chromatin of the floral repressor FLOWERING LOCUS C (FLC) and the sporophytically silenced floral repressor FWA.	[51]

Z. mays	[UniProt:O64411]	Polyamine oxidase	Catalyzes the oxidation of the secondary amino group of polyamines (spermine, spermidine and their acetyl derivatives)	[52, 53]
P. yoelii	[PlasmoDB:PY02606]; [UniProt:Q7RLD9]	amine oxidase, flavin-containing, putative	no data	[54]
P. chabaudi	[PlasmoDB:PCHAS_122900] UniProt (fragments): [UniProt:Q4Y046] [UniProt:Q4Y791] [UniProt:Q4X770] [UniProt:Q4XR08] [UniProt:Q4Y2X2] [UniProt:Q4Y406] [UniProt:Q4XKT6] [UniProt:Q4XRU1]	conserved Plasmodium protein, unknown function	no data	[12]
P. berghei	[PlasmoDB:PBANKA_122830] UniProt (fragments): [UniProt:Q4YHN0] [UniProt:Q4YNP6] [UniProt:Q4YV76] [UniProt:Q4YJV5] [UniProt:Q4YV75]	conserved Plasmodium protein, unknown function	no data	[12]
P. falciparum	[PlasmoDB:MAL8P1.154]/ [PlasmoDB:PF3D7_0801900] [UniProt:C0H4R0]	conserved Plasmodium protein, unknown function	no data	[55, 56]
P. vivax	[PlasmoDB:PVX_084610] [UniProt:A5K0L5]	hypothetical protein, conserved	no data	[57]
P. knowlesi	[PlasmoDB:PKH_131060]; [UniProt:B3LBA4]	lysine-specific histone demethylase 1, putative (LSD1)	no data	[58]

P. falciparum	[PlasmoDB:PFL0575w]/ [PlasmoDB:PF3D7_1211600]; [UniProt:Q8I5T5]	lysine-specific histone demethylase 1, putative (LSD1)	no data	[55]
P. yoelii	[PlasmoDB:PY03791]; [UniProt ID:Q7RI39]	amine oxidase, flavin-containing, putative	no data	[54]
P. vivax	[PlasmoDB:PVX_093645]; [UniProt:A5KCJ6]	hypothetical protein, conserved	no data	[57]
P. knowlesi	[PlasmoDB:PKH_011650]; [UniProt:B3KZC8]	reductase, putative	no data	[58]
P. berghei	[PlasmoDB:PBANKA_061010]; [UniProt:Q4YUI0] [UniProt:Q4YUH9] [UniProt:Q4YYR1] [UniProt:Q4Z5M6] [UniProt:Q4YH02]	lysine-specific histone demethylase 1, putative (LSD1)	no data	[12]
P. chabaudi	[PlasmoDB:PCHAS_061180] UniProt (fragments): [UniProt:Q4X7D7] [UniProt:Q4XV87] [UniProt:Q4XA38] [UniProt:Q4XNH7]	lysine-specific histone demethylase 1, putative (LSD1)	no data	[12]

Table S2. JmjC-containing proteins used in this study and their specificity towards methylated histones

Organism	**Accesion number (UniProt/ NCBI/ PlasmoDB)**	**Protein name**	**Histone demethylation activity**	**Citation**
H. sapiens	[UniProt:A6N6J7]	JARID1C	H3K4	[70]
H. sapiens	[UniProt:Q6B0I6]	Lysine-specific demethylase 4D, KDM4D	H3K9me3/me2	[71]
H. sapiens	[UniProt:Q9UGL1]	KDM5B	H3K4me3/me2/ me1	[72-75]
H. sapiens	[UniProt:P29375]	KDM5A	H3K4me3/me2	[74, 76-79]
H. sapiens	[UniProt:Q9BY66]	KDM5D/ JARID1D	H3K4me3/me2	[79-81]

H. sapiens	[UniProt:Q9H3R0]	KDM4C	H3K9me3 H3K36me3	[71]
H. sapiens	[UniProt:O75164]	KDM4A	H3K9me3 H3K36me3	[71, 82]
H. sapiens	[UniProt:O94953]	KDM4B (JMJD2B)	H3K9me3	[71]
H. sapiens	[UniProt:Q92833]	Protein Jumonji (JARID2)	None. Regulator of histone methyltransferase complexes	[83]
H. sapiens	[UniProt:Q8NHM5]	KDM2B	H3K4me3 H3K36me2	[29, 84]
H. sapiens	[UniProt:Q9UPP1]	PHF8	H3K9me2/me1, H3K27me2, H4K20me1, possibly H3K36me2	[60-69]
H. sapiens	[UniProt:Q9Y2K7]	KDM2A	H3K36me2	[29, 85]
M. musculus	[UniProt:Q62240]	KDM5D	By similarity: H3K4me3/me2	[86, 87]
M. musculus	[UniProt:Q3UXZ9]	KDM5A	H3K4me3/me2	[74, 88]
M. musculus	[UniProt:Q62315]	Protein Jumonji (Jarid2)	None. Regulator of histone methyltransferase complexes	[83, 89-97]
S. cerevisiae	[UniProt:P40034]	JmjC domain-containing histone demethylation protein 1 (JHD1)	H3K36	[29]
S. cerevisiae	[UniProt:P47156]	Histone demethylase JHD2	H3K4me3	[98]
S. pombe	[UniProt:Q9UT79]	Multicopy suppressor of chk1 protein 1 (MSC1)	H3 deacetylase	[99]
S. pombe	[UniProt:O94606]	JmjC domain-containing protein 4 (Jmj4)	no data	(role in meiosis; [100])

S. pombe	[UniProt:O94603]	Putative JmjC domain-containing histone demethylation protein 1 (Jhd1)	By similarity - H3K36, but might lack demethylase activity	[27, 29, 101]
D. melano-gaster	[UniProt:Q9VMJ7]	Lysine-specific demethylase lid	H3K4me3	[81, 102-104]
D. melano-gaster	[UniProt:Q8SXP6]	Jarid2	no data	-
V. dahliae	[UniProt:G2XGM5]	Histone demethylase JARID1A	no data	-
A. gypseum	[UniProt:E4URU6]	Histone demethylase JARID1C	no data	-
G. lozoyensis	[UniProt:H0EUH0]	Putative Lid2 complex component lid2	no data	- Genome: [105]
P. tritici-repentis	[UniProt:B2WM33]	Histone demethylase JARID1D	no data	-
C. elegans	[UniProt:Q23541]	Lysine-specific demethylase rbr-2	H3K4me3/me2	[74]
A. thaliana	[UniProt:Q8GUI6]	Probable lysine-specific demethylase JMJ14	H3K4me3/me2/me1	[106-110]
D. rerio	[UniProt:Q1LVC2]	Protein Jumonji / jarid2b	By similarity: None. Regulator of histone methyltransferase complexes	-
X.tropicalis	[UniProt:XP_002932734]	predicted: protein Jumonji	no data	-
G.gallus	[UniProt:Q5F363]	Protein Jumonji (JARID2)	By similarity: None. Regulator of histone methyltransferase complexes.	-

A.melano-leuca	[UniProt:XP_002917654]	predicted: protein Jumonji-like	no data	-
T.castaneum	[UniProt:D6WIN6]	Putative uncharacter-ized protein	no data	Genome: [111, 112]
A.echinatior	[UniProt:F4W6V2]	Protein Jumonji	no data	Genome: [113]
B.terrestris	[UniProt:XP_003396635]	predicted: hypothetical protein LOC100643807	no data	-
P.humanus corporis	[UniProt:E0VZT5]	Jumonji/arid domain-containing protein, putative	no data	Genome: [114]
A. carolinesis	[UniProt:G1KC73]	Uncharacter-ized protein / Jarid2, predicted	no data	-
A. gambiae	[UniProt:Q7PP11]	AGAP005541-PA, predicted	no data	Genome: [115]
A. aegypti	[UniProt:Q16RF7]	Putative uncharacter-ized protein	no data	Genome: [116]
P. yoelii	[PlasmoDB:PY05603]; [UniProt:Q7RD23]	Hypothetical protein	no data	Genome: [54]
P. chabaudi	PlasmoDB ID: PCHAS_091820 UniProt ID: Q4X9H4	Conserved protein, unknown function	no data	Genome: [12]
P. berghei	[PlasmoDB:PBANKA_092610]; [UniProt:Q4YRC5] [UniProt:Q4YRC6]	Conserved protein, unknown function	no data	Genome: [12]
P. falciparum	[PlasmoDB:PF11_0230]/ [PlasmoDB:PF3D7_1122200]; [UniProt:Q8IIE4]	Conserved protein, unknown function	no data	Genome: [55, 117]
P. vivax	[PlasmoDB:PVX_091730]; [UniProt:A5K4N3]	Hypothetical protein, conserved	no data	Genome: [57]

P. knowlesi	[PlasmoDB:PKH_091960]; [UniProt:B3L525]	Conserved protein, unknown function	no data	Genome: [58]
P. falciparum	[PlasmoDB:MAL8P1.111]/ [PlasmoDB:PF3D7_0809900]; [UniProt:Q8IAT4]	JmjC domain containing protein (JmjC1)	no data	Genome: [55, 56]
P. knowlesi	[PlasmoDB:PKH_142870] [UniProt:B3LCB0]	JmjC domain containing protein	no data	Genome: [58]
P. vivax	[PlasmoDB:PVX_123283]; [UniProt:A5JZT0]	Hypothetical protein	no data	Genome: [57]

The analysis of SET domain containing proteins from different species of Plasmodium revealed that they are more closely related to their orthologs from other *Plasmodia* than to other paralogs within a given species (Fig. 3). This means that diversification of SET-domain containing proteins preceded the diversification of *Plasmodium* species. We did not observe duplications of these proteins. SET-domains in Plasmodium are polyphyletic and form eight distinct classes. We also noticed that the orthologs of a protein [PlasmoDB: PF3D7_1133400] seemed to be deleted in different genomes of rodent *Plasmodium*, although at the time of a screen of Frech and Chen [36] it was overlooked.

For the basic prediction of histone lysine specificity of identified proteins we constructed an additional phylogenetic tree of their SET domains (Fig. 4). The tree included SET domains from vast variety eukaryotic taxonomic groups. The phylogeny thus included seven *P. falciparum* SET domains (in red) and showed that these domains represent different lineages of SET domains found across the eukaryotes. Interestingly, the SET domains of two proteins [PlasmoDB: PF3D7_1322100] and [PlasmoDB: PF3D7_0629700] appeared to be more closely related to mammalian SET domains than to the other single cell eukaryotes, such as yeast. The general domain composition in these proteins correlated closely with the mammalian domain composition. On the basis of the phylogenetic analysis (Fig. 4), we suggested specificity for six out of eight types of SET domain proteins.

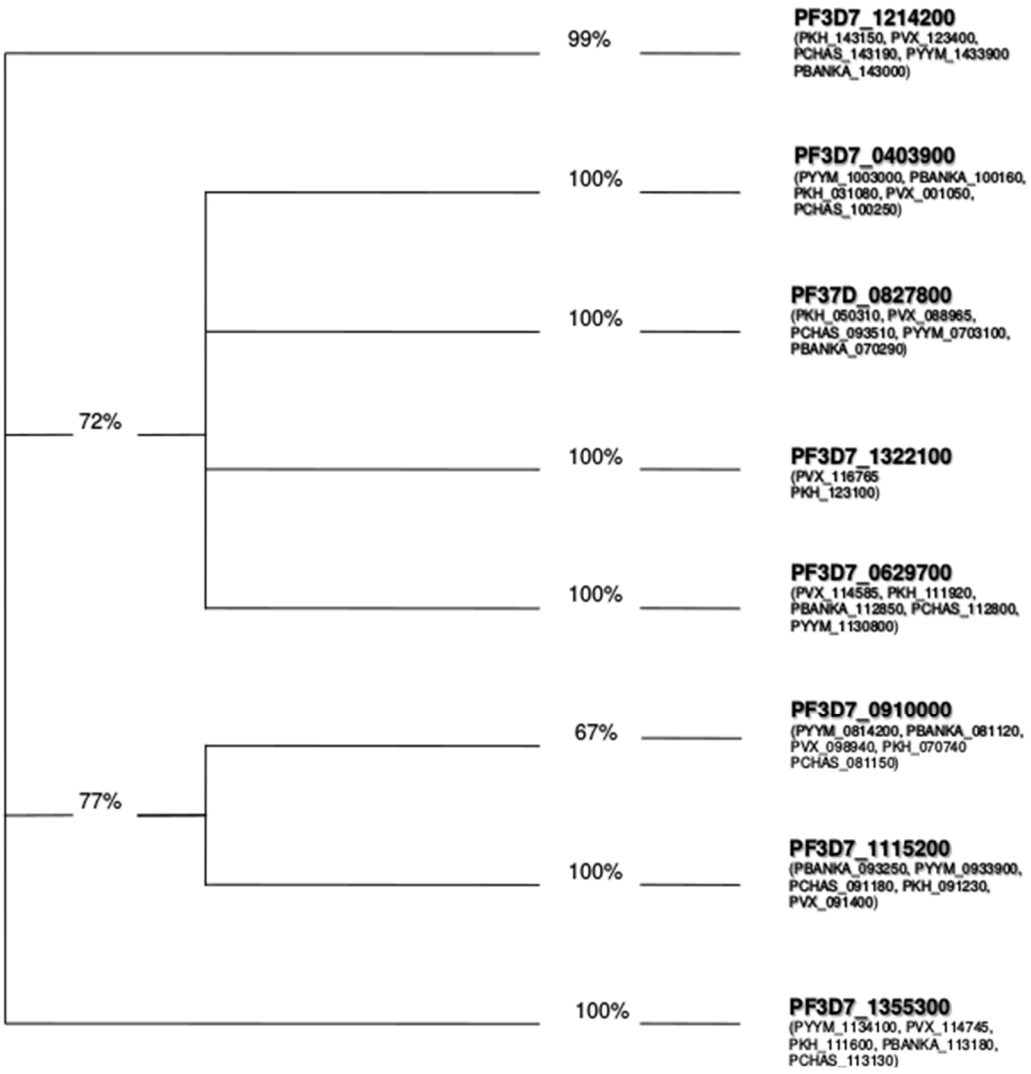

Fig. 3. Consensus phylogenetic parsimony tree of SET domain proteins of *Plasmodium* species.

Calculated using PHYLIP software with 1000 bootstrap replications. Protein alignment was edited and columns containing gaps were removed. There are no orthologs of [PlasmoDB:PF3D7_1322100] in rodent malaria parasites *Plasmodium berghei, Plasmodium chabaudi* and *Plasmodium yoelii.*

Fig. 4. Consensus phylogenetic parsimony tree of SET domain proteins of *Plasmodium* parasites and other eukaryotic organisms.

The tree was calculated using PHYLIP software with 1000 bootstrap replications. Protein alignment was edited and columns containing gaps were removed. The evolutionary distances were assessed using Phylip 3.67, Kitsch.

In order to verify predicted lysine methylation specificity of the [PlasmoDB:PF3D7_1322100] gene protein product (a protein missing in rodent but present in primate Plasmodia), we compared its amino acid

sequence with the known crystal structure of the H3K9 tri-methyltransferase DIM5 [UniProt: Q8X225], [PDB:1peg]. DIM-5 is a SUV39-type histone H3 Lys9 methyltransferase that is essential for DNA methylation in *N. crassa* [39]. The alignment of DIM5 with [PlasmoDB:PF3D7_1322100] indicated that this *Plasmodium* protein contains counterparts of key residues from the active center identified in DIM5. Specifically, [PlasmoDB:PF3D7_1322100] protein conserved Tyr178 of DIM5 [PDB:1peg] (Y191[UniProt: Q8X225]) and Tyr283 (Y296[UniProt: Q8X225]) (Fig. 5, highlighted in red), which are essential for DIM-5 catalytic activity. Zhang et al. [40] demonstrated that the 281 residue in DIM5 is crucial for determining the specificity of enzyme towards mono, di or tri-methylation. Namely, F281Y mutation changed the product specificity of DIM-5 from a tri-methylase to a mono- and di-methylasease without affecting its overall catalytic activity. [PlasmoDB:PF3D7_1322100] contains tyrosine residue at this position of Phe/Tyr-switch (Fig. 5, highlighted in yellow) corresponding to F281 in DIM5 [PDB:1peg] (i.e, Y191 in DIM5 [UniProt: Q8X225]) (Fig. 5, highlighted in green), which makes it likely to be a mono- or/and dimethylase.

```
DIM5          162VPLQIFRTKDRGWGVKCPVNIKRGQFVDRYLGEIITSEEADRRR----
MAL13P1.122  2119KDLEIKKTEKTGYGVFCKRDIKNGELICEYVGEVLGKREFEKRLEVYQ
NSD1_MOUSE   1840PDVEIFRTLQRGWGLRTKTDIKKGEFVNEYVGELIDEEECRARI----

DIM5          206AESTIARRKDVYLFALDKFSDPDSLDPLLAGQPLEVDGEYMSGPTRFIN
MAL13P1.122  2166EESKKTDMYNWYIIQINK--------------DVYIDSGKKGSISRFIN
NSD1_MOUSE   1884RYAQEHDITNFYMLTLDK--------------DRIIDAGPKGNYARFMN

DIM5          255HSCDPNMAIFARVGDHADKHIHDLALFAIKDIPKGTELTFDYVNGLTGL303
MAL13P1.122  2202HSCSPNSVSQKWIVR----GFYRIGIFALRDIPSGEEITYNYSYNFLFN2246
NSD1_MOUSE   1919HCCQPNCETQKWSVN----GDTRVGLFALSDIKAGTELTFNYNLECLGN1963
```

Fig. 5. The alignment of DIM5 [UniProt:Q8X225], NSD1 [UniProt:O88491] and MAL13P1.122 [PlasmoDB:PF3D7_1322100] proteins.

The alignment shows that the plasmodial protein is similar to NSD1 protein and it could be a histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific, like NSD1 proteins. Tyrosines corresponding to Y191 in DIM5 and Y294 are highlighted in red. Highlighted in green are phenylalanines corresponding to F281 in DIM5. The corresponding site in MAL13P1.122 is highlighted in yellow; it contains tyrosine which makes it likely to be mono- or/and dimethylase. Higlighted in blue are aspartate residues corresponding to D209 in DIM5. MAL13P1.122 lacks aspartate in

this position, which suggest it is not H3K9 specific but possibly specific to: H3K14, H3K36, H3K37 or/and the N-terminal Lys residues on the histone H4 tail.

The plasmodial protein does not contain aspartate residue in a position corresponding to Asp209 of DIM5 [PDB:1peg] (D222[UniProt: Q8X225]) (Fig. 5, highlighted in blue), which is involved in the recognition of the histone H3 tail, suggesting that [PlasmoDB:PF3D7_1322100] has a specificity that may not require the Asp-Ser interaction. The histone lysines that could be the potential targets for [PlasmoDB:PF3D7_1322100] are therefore those that lack a flanking Ser or Thr residue, like K14, K36, and K37 on histone H3, and the N-terminal Lys residues on the histone H4 tail.

The protein coded by [PlasmoDB:PF3D7_1322100] is annotated as containing PHD and SET domains. However, its homologous SET-domain proteins contain the cysteine rich AWS and PostSET domains, as well as the nuclear protein PWWP motif, none of which have been predicted in [PlasmoDB:PF3D7_1322100]. Still, a closer sequence investigation revealed that the [PlasmoDB:PF3D7_1322100] is indeed enriched in cysteines in the areas surrounding the SET domain, and appears to have sequences very similar to the consensus for the AWS and PostSET domains.

Domains surrounding the SET domain have been found to play structural roles important for the proper function of methyltransferase enzymatic activity. The AWS or PreSET domain has a structural function and holds together the SET domain through interactions between zinc ions and the conserved cysteines in this domain. The PostSET domain is important in forming the interaction with the cofactor and histone tails [41]. The presence of these domains in [PlasmoDB:PF3D7_1322100] supports the prediction of its functionality.

Altogether, on the basis of the above analyses we suggest that the protein product of [PlasmoDB:PF3D7_1322100] conforms to the expectation of being a monomethylase of H4K20 and H3K36.

3. Discussion. In this study we used a bioinformatic approach to analyze potential differences in histone methylation machinery between primate and rodent *Plasmodium* species. The basic homology and domain searches based on accessible databases led to identification of one putative

histone methylase and one putative histone demethylase that is in present in primate *Plasmodium* species but lost in rodent *Plasmodium* species. We cross validated the presence/absence of genes of interest in PlasmoDB with NBCI RefSeq and EST database using PSI Blast. The two genes were not found in all three rodent *Plasmodium* species. Moreover, the genome view in PlasmoDB shows that they are missing in the synteny blocks, and they seem to be just simply deleted.

We constructed the phylogenetic trees for homologs of both the aminooxidase and the JmjC domain containing putative histone demethylases in *Plasmodium* and found that there was no closely related paralog of a [PlasmoDB: PF3D7_0809900] gene product in rodent *Plasmodium* species, while there were no deletions in aminooxidase domain containing plasmodial putative histone demethylases. Since there are no recent duplications in [PlasmoDB: PF3D7_0809900] gene orthologs, their deletion in rodent *Plasmodia* could be evolutionarily old. This observation may suggest an evolutionary change in regulation of histone methylation in rodent and primate *Plasmodia,* provided that the orthologs of [PlasmoDB: PF3D7_0809900] retained their enzymatic or regulatory activity.

Accordingly, our analysis of putative SET-domain containing histone methylases of rodent and primate *Plasmodia* showed that there was one important deletion in rodent *Plasmodia* genomes, as they lack orthologs to *P. falciparum* putative histone methylase [PlasmoDB: PF3D7_1322100]. We found that there were eight types of putative histone SET-domain methyltransferases in primate *Plasmodia*, whereas in rodent *Plasmodia* only seven types were present. This analysis adds 3 more groups to the previously described 5 groups of SET-containing proteins in *Plasmodium* [42]. Each examined genome contained no more than one copy of each type of histone methyltransferase. Phylogenetic analysis indicated that all SET-domain containing proteins diverged before *Plasmodium* speciation, so similarly to the deletion of putative histone demethylase [PlasmoDB: PF3D7_0809900], the deletion of orthologs of the putative histone methylase [PlasmoDB:PF3D7_1322100] is also evolutionarily old.

The comparative analysis of the putative histone methylase lost in rodent *Plasmodia* [PlasmoDB:PF3D7_1322100] protein sequence with the

known crystal structure of the H3K9 tri-methyltransferase DIM5 indicated that [PlasmoDB:PF3D7_1322100] is a putative mono- or di-methylase of H3K36 and H4K20. At the same time, Cui et al. [42] suggested that the protein being a product of [PlasmoDB: PF3D7_0809900] gene could be a H3K36/ H3K9 demethylase, based on a comparison with JHDM3A/JHDM2A (KDM4A/KDM3A). These joint observations could mean that the whole module of proteins regulating methylation status of H3K36 was lost in rodent *Plasmodia* early on in evolution.

Parallel to the approach of Cui et al. [42] our phylogenetic analysis was based on isolated JmjC domains. In their study, Cui et al. [42] used a limited number of non-plasmodial proteins as references, mainly *S. cerevisiae* (5 proteins) and human (2 proteins). We used a wider range of JmjC domains. Our analysis showed that one of the closest homologs of [PlasmoDB: PF3D7_0809900] were not only H3K36 demethylase KDM4A shown before by Cui et al. [42] but also the Jumonji homologs. Jumonji has no demethylase activity, but is indispensable for the methylation status regulation of H3K9 and H3K27. Interestingly, it has been shown that plasmodial *var* genes undergo epigenetic regulation by methylation status of H3K9 [8]. Cui and coworkers [42] noticed the homology of putative plasmodial demethylases to Jumonji proteins, but they decided that the group including a [PlasmoDB: PF3D7_0809900] gene product more closely resembles the JHDM3A/2A family of proteins because of the presence of additional protein domains, especially JmjN [43]. Although in many JmjC proteins the JmjC domain is sufficient for demethylase activity [33] and JmjN domain seems unneccessary for catalysis as viewed from the protein structure [34], it is nevertheless required for enzymatic activity, at least in JHDM3/JMJD2A [44]. It therefore seems most likely that the [PlasmoDB: PF3D7_0809900] protein could be a H3K36 demethylase. Together with the observation that a putative H3K36 methylase orthologous to [PlasmoDB:PF3D7_1322100] is also missing in rodent *Plasmodia,* this would mean that the whole regulatory module of histone methylation was lost early on in the evolution of rodent *Plasmodia.*

4. Conclusions. We have carried out bioinformatics analyses of an important component of the global mechanism of epigenetic regulation of gene expression in *Plasmodium* species. Namely, we interrogated the repertoire of

enzymes involved in histone methylation; methyl transferases and demetylases. Crucially we found that one of the element represented by JmjC type of demethylase and the SET-type methylase is absent in the current published versions of the rodent *Plasmodium* genomes. Subsequently by interrogating sequence alignments between the *Plasmodium* proteins and their orthologues in other eukaryotic systems, we assessed putative substrate specificities of the existing enzymes. Our observations led us to a conclusion that the rodent *Plasmodium* species may have evolved different system of epigenetic regulation of gene expression.

5. Methods

5.1. Putative histone demethylases homology search. Based on the list of human histone demethylases [45] we found plasmodial homologs of these proteins using NCBI Protein (`http://www.ncbi.nlm.nih.gov/protein`), UniProt [46] and PlasmoDB [47] databases. We used High Throughput Sequence Annotation Service HT-SAS [48] to find additional homologs of known histone demethylases in the UniProt database [46] We also used an isolated JmjC domain model derived from Pfam [49] to search the PlasmoDB for other possible JmjC domain containg proteins. PlasmoDB was used to verify if the identified plasmodial proteins were orthologous. All proteins were investigated with Pfam homology searches for the existence of key domains: aminooxidase or JmjC. In every alignment, the e value was set smaller than 0.0001. The presence or absence of particular JmjC-domain containing proteins in *Plasmodium* was in addition verified with NCBI PSI-Blast. Out of 32 tested human histone demethylase and putative demethylase protein sequences, 14 had orthologs in *Plasmodium* identified by homology searches. Plasmodial proteins that contained aminooxidase, JmjC or cupin domain (a clan that also contains JmjC domain) were subject to further analysis.

5.2. SET-domain containing proteins homology search. The plasmodial protein sequences were derived from the PlasmoDB database and cross-checked with other databases (UniProt, EST NCBI). The known SET domain containing proteins within the base were identified using the plasmoDB web server (`http://plasmodb.org`). Additional proteins containing SET domains were identified with the use HMMER software. Two hidden

Markov profiles of the SET domain were obtained from the Pfam server: one Markov profile was configured for recognition of local alignments and the other for global alignment. PlasmoDB was searched with the SET profile calibrated for global alignments. All sequences of putative SET domain containing proteins were next compared with the whole database of hidden Markov profiles of all known Pfam domains. Low complexity sequences of *Plasmodium* proteins were masked. Only proteins identified by the profile configured for recognition of global alignments with an e-value smaller than 1 were included for this analysis (50 sequences in total). Sequences of SET domains were extracted from these sequences. The length of these SET domain sequences was determined using HMM search with hidden Markov models configured for a recognition of local alignments. We excluded from the analysis the proteins which Pfam identified as SET domains were too short to be reliably aligned (namely, the orthologs of [PlasmoDB:PF3D7_1221000] and [PlasmoDB:PF3D7_1304600]). We added to the analysis 5 more protein sequences that were not identified by HMMER, but found as orthologs of other putative SET containing methyltransferases within PlasmoDB, (namely: [PlasmoDB:PBANKA_093250, PCHAS_091180, PF3D7_0910000, PY00784, PY02023]. Finally we obtained a set of 45 proteins for a phylogenetic analysis.

5.3. Phylogenetic analysis. The protein domains (aminooxidase, cupin, JmjC, SET) were extracted from protein sequences using Pfam (`http://pfam.sanger.ac.uk`). The protein and domains alignments were calculated using MUSCLE [50]. For the histone demethylases tree construction (Fig. 1 and 2), and for the putative histone methyltransferases in *Plasmodium* tree construction (Fig. 3), the alignments were edited in Jalview [51], where the gaps were removed. For the tree covering *P. falciparum* SET domains from putative histone methyltransferases and SET domains from different eukaryotic homologs (Fig. 4), the SET containing proteins were aligned with a Pfam profile of SET domains by HMMalign software. Gaps were not removed from the alignments due to technical reasons (short sequences). All phylogenetic trees were constructed using Phylip software [52] with protein phylogeny calculated using the ProtPars algorithm with a bootstrap value of 100 and verified with maximum likelihood trees calculated using PhyML with a bootstrap value of 100 [53]. The evolutionary distances were assessed using Fitch-

Margoliash and Least Squares Methods with Evolutionary Clock (Phylip 3.67: Kitsch), basing on a distance matrix computed from aligned protein sequences using Phylip 3.67 [52]: Protdist.

Competing interests. The authors declare that they have no competing interests.

Authors' contributions. MP, PZ and SK performed research and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements. The authors acknowledge Prof. DL Hartl for his advice and thank Marna S. Costanzo for her input into the study. Work at the Institute of Biochemistry and Biophysics was supported by a grant from the National Science Centre of Poland, number 2014/13/B/NZ8/04719 for Szymon Kaczanowski.

REFERENCES

[1] World Health Organization: WHO World Malaria Report 2010. Edited by World Health Organization, 2011.

[2] Lee K. S., P. C. Divis, S. K. Zakaria, A. Matusop, R. A. Julin, D. J. Conway, J. Cox-Singh, B. Singh. Plasmodium knowlesi: reservoir hosts and tracking the emergence in humans and macaques. *PLoS Pathog*, **7** (2011), No 4, e1002015.

[3] Bozdech Z., J. Zhu, M. P. Joachimiak, F. E. Cohen, B. Pulliam, J. L. DeRisi. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. *Genome Biol*, **4** (2003), No 2, R9.

[4] Le Roch K. G., Y. Zhou, P. L. Blair, M. Grainger, J. K. Moch, J. D. Haynes, De La Vega P., A. A. Holder, S. Batalov, D. J. Carucci, E. A. Winzeler. Discovery of gene function by

expression profiling of the malaria parasite life cycle. *Science*, **301** (2003), 5639, 1503–1508.

[5] Cunningham D., J. Lawton, W. Jarra, P. Preiser, J. Langhorne. The pir multigene family of Plasmodium: antigenic variation and beyond. *Mol Biochem Parasitol*, **170** (2010) No 2, 65–73.

[6] Chookajorn T., R. Dzikowski, M. Frank, F. Li, A. Z. Jiwani, D. L. Hartl, K. W. Deitsch. Epigenetic memory at malaria virulence genes. *Proc Natl Acad Sci USA*, **104** (2007), No 3, 899–902.

[7] Duraisingh M. T., T. S. Voss, A. J. Marty, M. F. Duffy, R. T. Good, J. K. Thompson, L. H. Freitas-Junior, A. Scherf, B. S. Crabb, A. F. Cowman. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. *Cell*, **121** (2005), No 1, 13–24.

[8] Freitas-Junior L. H., Hernandez-Rivas R., S. A. Ralph, D. Montiel-Condado, O. K. Ruvalcaba-Salazar, A. P. Rojas-Meza, L. Mâncio-Silva, R. J. Leal-Silvestre, A. M. Gontijo, S. Shorte, A. Scherf. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*, **121** (2005), No 1, 25–36.

[9] Perez-Toledo K., Rojas-Meza A. P., Mancio-Silva L., Hernández-Cuevas N. A., D. M. Delgadillo, M. Vargas, S. Martínez-Calvillo, A. Scherf, R. Hernandez-Rivas. Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res*, **37** (2009), No 8, 2596–2606.

[10] Tonkin C. J., C. K. Carret, M. T. Duraisingh, T. S. Voss, S. A. Ralph, M. Hommel, M. F. Duffy, L. M. Silva, A. Scherf, A. Ivens, T. P. Speed, J. G. Beeson, A. F. Cowman. Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum. *PLoS Biol*, **7** (2009), No 4, e84.

[11] Volz J. C., R. Bártfai, M. Petter, C. Langer, G. A. Josling, T. Tsuboi, F. Schwach, J. Baum, J. C. Rayner, H. G. Stunnenberg, M. Duffy, A. F. Cowman. PfSET10, a Plasmodium falciparum methyltransferase, maintains the active var gene in a poised state during parasite division. *Cell Host Microbe*, **11** (2012), No 1, 7–18.

[12] Hall N., M. Karras, J. D. Raine, J. M. Carlton, T. W. A. Kooij, M. Berriman, L. Florens, C. S. Janssen, A. Pain, G. K. Christophides, K. James, K. Rutherford, B. Harris, D. Harris, C. Churcher, M. A. Quail, D. Ormond, J. Doggett, H. E. Trueman, J. Mendoza, S. L. Bidwell, M.-A. Rajandream, D. J. Carucci, J. R. Yates III, F. C. Kafatos, C. J. Janse, B. Barrell, C. M. R. Turner, A. P. Waters, R. E. Sinden. A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science*, **307** (2005), No 5706, 82–86.

[13] Kouzarides T. Chromatin modifications and their function. *Cell*, **128** (2007), No 4, 693–705.

[14] Lacoste N., R. T. Utley, J. M. Hunter, G. G. Poirier, J. Cote. Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. *J Biol Chem*, **277** (2002), No 34, 30421–30424

[15] Van Leeuwen F., P. R. Gafken, D. E. Gottschling. Dot1p Modulates Silencing in Yeast by Methylation of the Nucleosome Core. *Cell*, **109** (2002), 745–756.

[16] Feng Q., H. Wang, H. H. Ng, Erdjument-Bromage H., P. Tempst, K. Struhl, Y. Zhang. Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain. *Curr. Biol.,* **12** (2002), 1052–1058.

[17] Ng H. H., Q. Feng, H. Wang, H. Erdjument-Bromage, P. Tempst, Y. Zhang, K. Struhl. Lysine methylation within the globular domain of

histone H3 by Dot1 is important for telomeric silencing and Sir protein association. *Genes Dev.*, **16** (2002), 1518–1527.

[18] Bedford M. T., S. G. Clarke. Protein arginine methylation in mammals: who, what, and why. *Mol Cell*, **33** (2009), No 1, 1–13.

[19] Mosammaparast N., Y. Shi. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu Rev Biochem*, **79** (2010), 155–179.

[20] Wagner E. J., P. B. Carpenter. Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol*, **13** (2012), No 2, 115–126.

[21] Ringrose L., H. Ehret, R. Paro. Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. *Mol Cell*, **16** (2004), No 4, 641–653.

[22] Peters A. H., D. Schoeler. Methylation of histones: playing memory with DNA. *Curr Opin Cell Biol*, **17**(2005), No 2, 230–238.

[23] Tamada Y., J. Y. Yun, S. C. Woo, R. M. Amasino. ARABIDOPSIS TRITHORAX-RELATED7 is required for methylation of lysine 4 of histone H3 and for transcriptional activation of FLOWERING LOCUS C. *Plant Cell*, **21** (2009), No 10, 3257–3269.

[24] Jacob Y., S. Feng, C. A. LeBlanc, Y. V. Bernatavichute, H. Stroud, S. Cokus, L. M. Johnson, M. Pellegrini, S. E. Jacobsen, S. D. Michaels. ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol*, **16** (2009), No 7, 763–768.

[25] Cazzonelli C. I., T. Millar, E. J. Finnegan, B. J. Pogson. Promoting gene expression in plants by permissive histone lysine methylation. *Plant Signal Behav*, **4**(2009), No 6, 484–488.

[26] Margueron R., D. Reinberg. The Polycomb complex PRC2 and its mark in life. *Nature*, **469** (2011), No 7330, 343–349.

[27] Shi Y., F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, **119** (2004), No 7, 941–953.

[28] Metzger E., M. Wissmann, N. Yin, J. M. Müller, R. Schneider, A. H. Peters, T. Günther, R. Buettner, R. Schüle. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature*, **437** (2005), No 7057, 436–439.

[29] Karytinos A., F. Forneris, A. Profumo, G. Ciossani, E. Battaglioli, C. Binda, A. Mattevi. A novel mammalian flavin-dependent histone demethylase. *J. Biol. Chem.*, **284** (2009), No 26, 17775–17782.

[30] Ciccone D. N., H. Su, S. Hevi, F. Gay, H. Lei, J. Bajko, G. Xu, E. Li, T. Chen. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature*, **461** (2009), No 7262, 415–418.

[31] Trewick S. C., McLaughlin P. J., R. C. Allshire. Methylation: lost in hydroxylation? *EMBO Rep*, **6** (2005), No 4, 315–320.

[32] Takeuchi T., Y. Yamazaki, Y. Katoh-Fukui, R. Tsuchiya, S. Kondo, J. Motoyama, T. Higashinakagawa. Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev*, **9** (1995), No 10, 1211–1222.

[33] Tsukada Y., J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers, P. Tempst, Y. Zhang. Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, **439** (2006), No 7078, 811–816.

[34] Klose R. J., E. M. Kallin, Y. Zhang. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet*, **7** (2006), No 9, 715–727.

[35] Seward D. J., G. Cubberley, S. Kim, M. Schonewald, L. Zhang, B. Tripet, D. L. Bentley. Demethylation of trimethylated histone H3 Lys4 in vivo by JARID1 JmjC proteins. *Nat Struct Mol Biol*, **14** (2007), No 3, 240–242.

[36] Frech C., N. Chen. Genome comparison of human and non-human malaria parasites reveals species subset-specific genes potentially linked to human disease. *PLoS Comput Biol*, **7** (2011), No 12, e1002320.

[37] Korbel J. O., T. Doerks, L. J. Jensen, C. Perez-Iratxeta, S. Kaczanowski, S. D. Hooper, M. A. Andrade, P. Bork. Systematic association of genes to phenotypes by genome and literature mining. *PLoS Biol.*, **3** (2005) No 5, e134.

[38] Yamada T., M. Kanehisa, S. Goto. Extraction of phylogenetic network modules from the metabolic network. *BMC Bioinformatics*, **7** (2006), 130.

[39] Tamaru H., E. U. Selker. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature*, **414** (2001), No 6861, 277–283.

[40] Zhang X., Z. Yang, S. I. Khan, J. R. Horton, H. Tamaru, E. U. Selker, X. Cheng. Structural basis for the product specificity of histone lysine methyltransferases. *Mol Cell*, **12** (2003), No 1, 177–185.

[41] Xiao B., C. Jing, G. Kelly, P. A. Walker, F. W. Muskett, T. A. Frenkiel, S. R. Martin, K. Sarma, D. Reinberg, S. J. Gamblin, J. R. Wilson. Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev*, **19** (2005), No 12, 1444–1454.

[42] Cui L., Q. Fan, J. Miao. Histone lysine methyltransferases and demethylases in Plasmodium falciparum. *Int J Parasitol*, **38** (2008), No 10, 1083–1097.

[43] Klose R. J., Q. Yan, Z. Tothova, K. Yamane, H. Erdjument-Bromage, P. Tempst, D. G. Gilliland, Y. Zhang, W. G. Kaelin. The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell*, **128** (2007), No 5, 889–900.

[44] Klose R. J., K. Yamane, Y. Bae, D. Zhang, H. Erdjument-Bromage, P. Tempst, J. Wong, Y. Zhang. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature*, **442** (2006) No 7100, 312–316.

[45] Pedersen M. T., K. Helin. Histone demethylases in development and disease. *Trends Cell Biol*, **20** (2010), No 11, 662–671.

[46] UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res*, **40** (2012), Database issue, D71–75.

[47] Aurrecoechea C., J. Brestelli, B. P. Brunk, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O. S. Harb, M. Heiges, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, J. A. Miller, V. Nayak, C. Pennington, D. F. Pinney, D. S. Roos, C. Ross, C. J. Stoeckert Jr., C. Treatman, H. Wang. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res*, **37** (2009), Database issue, D539–543.

[48] Kaczanowski S., P. Siedlecki, P. Zielenkiewicz. The High Throughput Sequence Annotation Service (H. T.-SAS)—the shortcut from sequence to true Medline words. *BMC Bioinformatics*, **10** (2009), 148.

[49] Punta M., P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. L. Sonnhammer, S. R. Eddy, A. Bateman, R. D. Finn. The Pfam protein families database. *Nucleic Acids Res*, **40** (2012), Database issue, D290–301.

[50] Edgar R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, **32** (2004), No 5, 1792–1797.

[51] Waterhouse A. M., J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25** (2009), No 9, 1189–1191.

[52] Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, 2004.

[53] Guindon S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, **59** (2010), No 3, 307–321.

[54] Han M. V., C. M. Zmasek. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, **10** (2009), 356.

Malgorzata Perycz
e-mail: mperycz@ibb.waw.pl
Piotr Zielenkiwiecz
e-mail: piotr@ibb.waw.pl
Szymon Kaczanowski
e-mail: szymon@ibb.waw.pl
Institute of Biochemistry and Biophysics
Polish Academy of Sciences
02-106 Warsaw, Poland

Received November 10, 2017
Final Accepted February 28, 2018