DOI 10.18699/SBB-2022-100

Classification of prokaryotic DNA methyltransferases by sequence similarity of the catalytic domain

Volobueva M.[1, 2]*, Rusinov I.[1], Karyagina A.[1, 3, 4], Spirin S.[1, 2, 5], Alexeevski A.[1, 2, 5]

[1] *Belozersky Institute of Physical and Chemical Biology, Lomonosov Moscow State University, Moscow, Russia*

[2] *Faculty of Bioengineering and Bioinformatics, Lomonosov MSU, Moscow, Russia*

[3] *Gamaleya National Research Center for Epidemiology and Microbiology, Moscow, Russia*

[4] *Institute of Agricultural Biotechnology, RAS, Moscow, Russia*

[5] *Scientific Research Institute for System Analysis, RAS, Moscow, Russia*

** mashila6799@gmail.com*

Key words: DNA methyltransferase, classification, restriction-modification system

Motivation and Aim: DNA methylation is widely used by prokaryotic organisms in functioning of restriction-modification systems, and also for genome repair, regulation of gene expression and many other purposes. DNA methyltransferases (MTases) have been classified by methylation type (m^5C, m^6N, and m^4C) [1], and the order of conservative motifs in amino acid (aa) sequences [2]. In REBASE (http://rebase.neb.com) MTases are also classified by the R-M system types (Type I, Type II, Type III, and their subtypes). Universal aa sequence similarity-based classification of DNA MTases is still missing, but such classification could be useful for computational detection and annotation of MTase genes in newly sequenced genomes and metagenomes. The main aim of this work is the classification of all known prokaryotic MTases by sequence similarity of the catalytic domain.

Methods and Algorithms: All currently available 230873 annotated MTase sequences were obtained from REBASE (ver. 108), which are components of restrictionmodification systems, as well as solitary MTases. We use HMM profiles from Pfam (https://pfam.xfam.org, ver. 34.0) and SUPFAM (https://supfam.org/SUPERFAMILY, ver. 1.75) to determine domain architectures of the MTases with hmmsearch (e-value < 0.01). HMM profiles that correspond to catalytic MT domains were selected.

Results: We propose a new classification of MTases based on HMM profiles from the Pfam and SUPFAM databases. The classification consists of two levels. The first level is the classification according to the sequence similarity and proposed homology of catalytic domains of these enzymes. At the second level, a class of the first level is divided into groups corresponding to domain architectures, i.e. sequences of all Pfam domains in sequence. Thus, one group includes proteins that have similar sequences along their entire length. In total, we selected eight HMM profiles from the Pfam database and 12 HMM profiles from the SUPFAM database that locate the catalytic domain in the MTase sequence. On the basis of known 3D structures of MTase classes' representatives, we identified in HMM profiles regions that correspond to SAM-binding and catalytic subdomains and motifs. We compared and contrasted the HMM profiles of Pfam and SUPFAM catalytic domains (Table 1).

Table 1. Comparison of Pfam (rows) and SUPFAM (columns) profiles. The cells list the numbers of proteins with overlapping Pfam and SUPFAM profiles. The high enough values are marked in bold. Only sequences having both Pfam and SUPFAM hits are included

Family abbreviation	SF-DAM		SF-DCM			SF-N6_DNA			SF-TaqI	SF-TypeII		
Model	48856	52484	45633	46303	46923	51816	52618	54378	53087	36976	37952	45988
DNA_methylase	1	287	**4096**	**23667**	**13198**	2	21	214	37	0	3	64
Dam	0	3	5	60	8	0	0	0	0	1	0	0
Eco57I	0	0	1	112	111	4	49	**6086**	285	0	0	0
EcoRI_methylase	0	0	0	1	3	0	0	0	0	0	0	0
MT-A70	0	0	0	5	0	0	1	0	4	0	0	0
MethyltransfD12	**3533**	**31140**	49	181	57	0	3	0	0	0	0	14
N6_Mtase	0	0	2	16	9	**42947**	**208**	**7790**	52	0	0	0
N6_N4_Mtase	7	2	3	36	29	0	0	1	0	**3699**	**1807**	**39563**

Based on a comparison of overlaps of matches of HMM profiles containing full SAMbinding and catalytic subdomains, all 20 profiles were divided into the following groups: i) DNA_methylase, SF-DCM; ii) MethyltransfD12, SF-DAM; iii) N6_N4_Mtase, SFTypeII; iv) N6_Mtase, Eco57I, SF-N6_DNA, SF-TaqI; v) Dam; vi) MT-A70; vii) EcoRI_methylase. At least one of the above profiles is found in 90 % of MTases from REBASE.

Conclusion: The MTases were classified according to sequence similarity. The developed classification significantly fills the gaps of the previous classifications.

Acknowledgements: The study is supported by the Russian Science Foundation (grant No. 21-14-00135).

References

1. Wilson G.G., Murray N.E. Restriction and modification systems. *Annu Rev Genet*. 1991;25:585-627.
2. Bujnicki J.M. Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol Biol*. 2002;2:3.