DOI 10.18699/SBB-2022-083

Study of distribution and evolution of restriction-modification systems with restriction endonuclease of the AlwI family

Makarikova O.[1]*, Popova A.[2], Alexeevski A.[3, 4]

[1] *Lomonosov Moscow State University, Moscow, Russia*

[2] *Moscow Polytechnical University, Moscow, Russia*

[3] *A.N. Belozersky Research Institute, Lomonosov Moscow State University, Moscow, Russia*

[4] *FSI FSC SRISA RAS, Moscow, Russia*

* *makarikovaolga82@gmail.com*

Key words: restriction-modification, Pfam domains, phylogeny

Motivation and Aim: Restriction-modification (R-M) is the defense system of prokaryotes against phage and plasmid infection. The system has two enzymatic activities: restriction endonuclease (RE) and methyltransferase (MTase). The former hydrolyzes invaded DNA within or near their recognition site whereas the latter prevents self degradation of the cell genome by methylation of the recognition site. The system type II REs of the AlwI family represented in Pfam by the RE_AlwI domain are promised for bioengineering usage as nickases [1]. The aim of this work is to study the distribution and evolution (horizontal gene transfer) of R-M systems containing proteins of the family.

Methods and Algorithms: Pfam HMM-profile RE_AlwI was marked according to the structural domains (DNA-recognition, linker, catalytic) of the solved 3D structure of N.BspD6I nickase [1] (pdb code 2EWF) by aligning the sequence of it with the profile. New HMM-profile cat_AlwI for complete catalytic domain was created using HMMER3 on the base of a validated and non-redundant sample of 87 sequences of the catalytic domain only from REBASE (database containing information on all R-M systems) proteins with the RE_AlwI domain.

HMM-profiles RE_AlwI and cat_AlwI were compared by searching all proteins of 1089 complete prokaryotic proteomes from Uniprot database, containing at least one protein with the RE_AlwI Pfam domain, according database reference field in the entry annotation.

We have classified R-M systems as follows. A class is defined as a group of R-M systems with the same sets of catalytic domains of REs and MTases. All 818 R-M systems in REBASE having RE with the cat_AlwI domain were separated and classified. For the most represented class with one RE and two MTases of the MethyltransfD12 Pfam family genome coordinates of the genes of RE proteins found by the cat_AlwI profile in the abovementioned 1089 proteomes and MTases found in the same 1089 proteomes by the MethyltransfD12 profile were obtained. Proteins located in the same locus and the most likely associated with the same R-M system were selected.

Phylogenetic trees of 79 RE and 158 unique MTase sequences were built for further evolution analysis.

Results: The RE_AlwI domain profile was shown to cover only a small part of the catalytic domain. Therefore we have constructed the new cat_AlwI domain based only on the catalytic one. Hits of the cat_AlwI HMM-profile were found in 1005 REs with RE_AlwI domains according to the Uniprot entry annotation. We have shown that

84 sequences from 1089 complete prokaryotic proteomes containing RE_AlwI but not cat_AlwI domain are found by RE_AlwI profile due to the similarity of the DNA-binding and the linker domains rather than the catalytic one. Therefore the cat_AlwI profile proved to better represent the AlwI family and it has been used for further investigations.

Intriguingly R-M systems having RE with the cat_AlwI domain make up 10 classes while their corresponding methyltransferases belong to 3 different families. The class with one RE and two MTases of the MethyltransfD12 Pfam family turned out to be the most represented.

For class including 24 R-M systems with RE of AlwI family and two MTases of N6_N4_mtase family, phylogenetic tree of 48 MTases was shown to divide into two clades of equal size, so that one MTase belongs to one clade and the second from the same system to another. The uncommon separation of MTases tree turned out to occur in the most represented class for 158 MTase sequences as well. We are excited to see the differences in the MTase sequences that may be associated with non palindromic recognition sites which requires further investigation.

Conclusions: The new profile based on the catalytic domain of N.BspD6I nickase was shown to better represent the AlwI RE family. The results obtained in the phylogenetic analysis suggest that two MTases belonging to the one R-M system undergo divergent evolution. Differences observed in the MTase sequences may have occurred in response to emergence of non palindromic recognition sites.

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

References

1. Zheleznaya L.A. et al. Nicking endonucleases. *Biochemistry (Moscow).* 2009;74(13):1457-1466.