Identification of Novel Argonaute Proteins Using a Metagenomic Mining Approach

Lobna Ghonaim.
Graduate Program of Biotechnology, School of Sciences and Engineering, The American University in Cairo, New Cairo, Egypt

Ali H. A. Elbehery
Department of Microbiology and Immunology, Faculty of Pharmacy, University of Sadat City, Sadat City, Egypt

Ahmed Moustafa
Graduate Program of Biotechnology, School of Sciences and Engineering, The American University in Cairo, New Cairo, Egypt, Department of Biology, School of Sciences and Engineering, The American University in Cairo, New Cairo, Egypt

Abstract:

Synthetic biology combines biology, engineering, and computer sciences to design novel biological molecules and systems tailored for specific purposes. One of its most promising tools is gene editing, which enables precise modifications to an organism's genetic code. On the other hand, using metagenomics serves as a powerful tool that unlocks the broad genetic potential found in uncultured microbial communities. Exploring the untapped genetic diversity of uncultured microbial communities helps identify novel functional proteins with properties and makes best use of the potential of diverse microbial ecosystems.

We developed and employed a metagenomic-based approach to mine for prokaryotic argonaute proteins (pAGOs), an potential gene editing machinery encoded in bacterial and archaeal genomes. Our workflow involved strict quality control, sequence assembly, gene prediction, taxonomic classification, and annotation. We then followed manual validation to accurately identify key domains such as PIWI, MID, and PAZ. Our methodology allowed for the effective screening and identification of these proteins across metagenomes from 25 different ecosystems, ranging from island sediments to freshwater environments.

We analyzed 1,011 publicly available metagenomic datasets from which we identified 1,451 new putative pAGOs across these diverse environments by constructing a custom Hidden Markov Model (HMM) profile, highlighting their significance in various microbial ecosystems. A Markov clustering (MCL)-based approach revealed the presence of four distinct clusters of pAGOs characterized by unique domain architectures and functional specializations. These clusters demonstrate significant variability in the distribution of argonaute proteins across different biomes. Representative proteins from each cluster were further analyzed for their tertiary structure. Notably, our analysis shows the prevalence of argonaute proteins in environments with high microbial diversity and complex defense mechanisms, suggesting a critical role in microbial immunity.

Our work highlights the power of metagenomics to discover novel functional proteins. The findings contribute to a deeper understanding of RNA interference pathways and present promising applications for precise gene editing. Additionally, the study improves our knowledge of microbial gene regulation mechanisms and serves as a potential nucleus for various biotechnological applications, including therapeutics and diagnostics.

Keywords:

metagenomics, argonaute proteins, genetic editing, gene silencing, bioinformatics.