Accelerated Protein Evolution in Bacteria to Facilitate Environmental Adaptation: Efficient Bioinformatics Algorithms to Discover Diversity-Generating Retroelements (DGRs)

Ganesh, Anakha (School: James Logan High School)

Diversity generating retroelements (DGRs) are genetic elements that promote hyper mutations in specific regions of the genome. The targeted mutations of DGRs allow positive traits to manifest quicker than in traditional evolution. These mutations, predominantly in bacteria, enable species to adapt to environmental stress. Current DGR identification methods, such as myDGR, mis-annotate Group II introns (another region of the genome) as a DGR and can't be incorporated into larger bioinformatics workflows. The objective was to develop the automated portable software DGRpy to discover and cluster the template repeat (TR) and variable repeat (VR) pairs in a DGR with increased accuracy. DGRpy was integrated with the Prodigal gene prediction method (an unsupervised machine learning algorithm) and HMMER sequence and profile comparison method to identify potential DGR reverse transcriptase. The raw output was then parsed to extract smaller regions from the genome with potential DGRs. The sequences were split into overlapping windows applying the sliding window algorithm to ensure the entire TR/VR region was accurately identified. Sequence data was then algorithmically processed for the adenine mutations characteristic of DGRs. Finally, data clustering was applied to remove redundancies. Testing DGRpy with cyanobacterial genomes yielded the confusion matrix score: 82% accuracy, 90% sensitivity, 70% specificity, and 82% precision with minimal redundancy. DGRpy is being used in ongoing research to identify DGRs in Spartina grass microbes, a keystone species in its environment, to study its resilience to environmental stress. DGRpy automates DGR study to provide insights into biochemical processes necessary to survive climate change.

Awards Won:
Fourth Award of $500