CONSIDER FOR TALK

2024 SEA Symposium Abstract

University of Connecticut

Storrs CT

Corresponding Faculty Member: Johann Peter Gogarten (gogarten@uconn.edu)



Gabrielle F Stack

Intein Containing Genes in Actinophages From Subclusters A1 and C1

Gabrielle F Stack, Dehal H Dharani, Pamela Samuel, Najeed I Rahman, Jack R Coburn, Mahit Gupta, Sam A Gilden, Neil S Komma, Emily A Lawrence, Charlene Lee, Vincenzo J Marmo, Dharmik U Patel, Johann Peter Gogarten, Danielle Arsenault

Introduction: Inteins, aka protein introns, are self-splicing insertion sequences similar to introns; however, they only remove themselves from the host protein after translation. Most inteins consist of a self-splicing and a homing endonuclease (HEN) domain. The former minimizes harm to the invaded protein, the latter allows for invasion of previously uninvaded homologs. Most frequently the HEN is of the LAGLIDADG type, named after the amino acid sequence of a conserved motif. In case of mini inteins the HEN domain is absent. Inteins originate as molecular parasites infecting phages which in turn infect bacteria (which in turn may infect other organisms). Similar to Augustus De Morgan’s poem Siphonaptera (1872): “Great fleas have little fleas upon their backs to bite 'em, And little fleas have lesser fleas, ...”.   
Methods and Results: Genes containing putative inteins were identified using Position Specific Scoring Matrices calculated using known inteins as seeds.   
We identified seven putative intein containing genes in subcluster C1, and 4 in subcluster A1. We collected intein-containing and intein-free homologs of the identified genes.   
The presence of in insertion sequence was confirmed using alignments between intein-containing and intein-free homologs in blast searches, phamerator alignments, and Gepard dot plots. For each insertion sequence the identification as an intein was confirmed using HHPred, and structure predictions using alphafold. All predicted structures included a self-splicing domain that aligned well with the self-splicing domains of inteins whose structure had been determined by X-ray crystallography. In addition, all identified inteins had a domains that aligned to known HENs.   
Inteins comprise a large fraction of the encoded protein. Therefore, most intein-containing genes were assigned a Pham different from the intein-free homologs.   
We found inteins in well-characterized proteins such as terminases, portal proteins, helicases, and primases. In these instances, the intein-containing and intein-free homologs were assigned the same functions. However, we also found inteins in gene families that were not assigned a known function, and the intein-free and intein-containing homologs that were assigned different functions; e.g. HNH endonuclease versus nucleotidyltransferase; or HNH Endonuclease vs Cas 4 family exonuclease. While the intein contains an endonuclease domain, this domain does not contain the motif typical for HNH endonucleases, but rather the LAGLIDADG motif.   
Conclusion: Many different protein coding genes in phages from the A1 and C1 subcluster were invaded by inteins. In only a single gene annotation was the intein recognized as intein with LAGLIDADG HEN.