CONSIDER FOR TALK

2025 SEA Symposium Abstract

University of Connecticut

Storrs CT

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

Molecular Parasites Invade a Wide Range of Phage Genes and Clusters (A1, C1, E, EN, J)

A. North Ackley, Komla Amezouwoe, Kyle Fernand, Barry Held, Alexander Lee, Yaeyoung Min, Ivy Petrosky, Abhinav Wadhwa, Ryan Wong, Hanbo Zhang, Danielle Arsenault, Johann Peter Gogarten

*“Great fleas have little fleas upon their backs to bite 'em,
And little fleas have lesser fleas, and so ad infinitum.
And the great fleas themselves, in turn, have greater fleas to go on;
While these again have greater still, and greater still, and so on.”*
- Siphonaptera (1872) Augustus De Morgan

**Introduction**: Inteins are molecular parasites of phages, which infect bacteria, which infect other organisms. Mechanistically, inteins insert themselves within a gene, similar to introns. Unlike introns, which excise at the RNA level, inteins excise themselves post-translationally by protein splicing. Full inteins contain a self-splicing domain, allowing them to perform their seamless excision, and a central homing endonuclease (HE) domain. The HE bestows the intein with the ability to invade homologs of its host gene which have not yet been invaded, and to do so with super-Mendelian efficiency. Inteins engage in prolific horizontal transfer; however, once it has invaded most of a population selection for a functioning HE decreases. Thus, the HE begins to decay while the splicing domain remains intact, becoming a so called mini intein. While inteins are found across all domains of life, recent increased focus on inteins in phages has revealed their incredible abundance, diversity, and activity. In this work, ten phage insertion sequences identified in host genes including terminase large subunits, portal proteins, helicases, primases, and more which are suspected to be inteins are being investigated with computational approaches including homolog retrieval, protein structure prediction, and phylogenetic analysis.

**Methods**: A preliminary scan of PhagesDB sequences using intein-based position specific scoring matrices and an iterative search approach was conducted, retrieving many putative inteins in the database. Each student researcher is investigating one of these sequences to assess whether or not it is an intein. This begins with conducting searches against PhagesDB to retrieve insertion-free and insertion-containing homologs of their assigned gene. Then, a large multiple sequence alignment is generated from this collection of sequences. This alignment is used in conjunction with Phamerator maps to visualize the insertion. The host gene and insertion sequences can be extracted from this alignment, realigned on their own, and used as input with the phylogenetic reconstruction program IQ-TREE to generate unrooted phylogenies and dissect the distribution of the insertion across the represented phages. This process can reveal interesting dynamics such as cases of horizontal transfer. Additionally, the insertion alone can be extracted and used in databank searches in combination with AlphaFold3 to have sequence and structure-based evidence supporting or rejecting the insertion’s identity as an intein.