CONSIDER FOR TALK

7th Annual SEA-PHAGES Symposium Abstract

Brown University

Providence RI

Corresponding Faculty Member: Sarah Taylor (sarah\_taylor@brown.edu)



Julia Gross

Characterization of Palindrome Usage in Mycobacteriophage Genomes

Julia Gross, Kenyon Alexander, Brandon Le, Zachary Ricca Ricca, Hwai-Ray Tung, Benjamin Siranosian, Yang Zhou

Palindromes (DNA sequences with identical reverse complements) are thought to occur infrequently in bacteriophage genomes. They are often targets of Type II restriction enzymes, used by bacteria as part of restriction-modification systems. However, M. smegmatis does not code for any restriction enzymes and the Mycobacterium genus contains few restriction enzymes in general. We examined the presence of length-4 and length-6 palindromes in the genomes of all sequenced mycobacteriophages. Palindrome usage is varied across phage genomes, but we found interesting cases of both over and under-used palindromic sequences.   
  
In accordance with the idea of under-representation of palindromes, all cluster A4 phages lack the sequence ATAT and half of cluster A2 phages lack AATT. We tested the probability of this occurring in genomes with the same length and GC content but randomized nucleotide order. The probability of a randomized A4 genome not having an ATAT sequence is less than 10^-23. A selective pressure must be suppressing the occurrence of these palindromes. Similarly, some genomes do not have a single example of certain length-6 palindromes; these palindromes are frequently AT-rich. Surprisingly, there is no correlation between sequence length and the number of unused length-6 palindromes.   
  
Previous research at Brown highlighted the frequent usage of GATC and GGATCC palindromes in cluster B3 mycobacteriophages. We examined genes from B3 phages with strong homology to non-B3 cluster B phages. After alignment with BLAST, 73% of B3 GATC sites had mismatches or gaps in the comparison sequence. Most mismatch sites contained a single G to C transversion. These results argue that B3 phages commonly maintain a palindromic GATC site where related phages have NATC.  
  
In this project we sought to shed light on examples of both palindromic over and under-representation in hopes of elucidating what role palindrome frequencies play in phage biology and the evolutionary pressures driving outliers. Furthermore, investigating palindrome usage is an excellent way to teach the basics of computational biology, genome analysis, and data processing to students in the SEA-PHAGES program.