

1056-92-1902

Ming-Ying Leung*, 500 W. University Avenue, Department of Mathematical Sciences, El Paso, TX 79968-0514. *Poisson Approximations for Palindrome Distributions in DNA Viral Genomes.*

Palindromes are symmetrical words of DNA in the sense that they read exactly the same as their reverse complementary sequences. Representing the occurrences of palindromes in a DNA molecule as points on the unit interval, the scan statistics can be used to identify regions of unusually high concentration of palindromes. These regions have been associated with the replication origins on a few herpesviruses in previous studies. However, the use of scan statistics requires the assumption that the points representing the palindromes are independently and uniformly distributed on the unit interval. We provide a mathematical basis for this assumption by showing that in randomly generated DNA sequences, the occurrences of palindromes can be approximated by a Poisson process. Furthermore, we propose a compound Poisson approximation for the Palindrome Length Score (PLS) as a measure of palindrome concentration in fixed-length windows in viral DNA. As this compound Poisson distribution for PLS can be explicitly computed, it provides criteria for evaluation of statistical significance of the PLS. Windows with significantly higher PLS can then be identified as likely location of replication origins. This approach has been applied to predict replication origins for herpesvirus genomes. (Received September 22, 2009)