

1096-92-2320

Neal Williams* (nfwilli2@asu.edu). *De Bruijn Graph Analysis of Transposable Elements in DNA*.

The recent availability of inexpensive, short-read DNA sequencing data has created the need for algorithms that can efficiently assemble this data into contiguous genomic sequences. Many popular assemblers utilize a De Bruijn graph structure that represents each sequence of length k as a separate node in the assembly graph and connects nodes with sequences that overlap by $k - 1$ nucleotides with directed edges. This assembly problem is confounded by the existence of transposable elements (TEs), which are potentially highly-repetitive DNA sequences that are able to reinsert themselves into the genome at different loci. The TE profiles for eukaryotes can vary widely between species, which suggests a volatile evolutionary history that is only partially understood. We have implemented an algorithm that finds DNA sequences that align to the consensus sequence for known TE families, and then uses this data to construct De Bruijn assembly graphs which can be visualized to aid interpretation. In the talk, we will explore what insights this approach reveals about the complexities that TE families induce in complete assembly graphs, the effectiveness of long read data in reducing assembly graph complexity, and the TE variation observed across different tissues, organisms, and species. (Received September 17, 2013)