Detecting complexities in a scrambled genome through spacial graphs.

DNA rearrangement is a process found on both developmental and evolutionary scale. The process itself and the molecular shape at the time of the rearrangement can be modeled through 4-regular graphs. These graph models are illustrated through the rearrangement processes in a well studied ciliate species Oxytricha trifallax where DNA recombination is observed on a massive scale. Our studies show that there are two general patterns, reoccurring genome-wide, that describe over 90% of the Oxytricha’s scrambled genes. Further, gene segments that recombine during DNA rearrangement processes may be organized on the chromosome in a variety of ways. They can overlap, interleave or one may be a subsegment of another. We use colored directed graphs to represent contigs containing rearranged segments where edges represent recombining segment organization. Using graph properties we associate a point in a higher dimensional Euclidean space to each graph such that cluster formations and analysis can be performed with methods from topological data analysis. The analysis shows some emerging graph structures indicating that segments of a single gene can interleave, or even contain, all of the segments from several other genes in between its segments. (Received August 28, 2018)