

SDRAP: An Annotation Pipeline for Highly Scrambled Genomes

Jasper Braun^{1,*}, Nataša Jonoska¹, Rafik Neme^{2,3}, Laura Landweber²

¹Department of Mathematics & Statistics, University of South Florida, Tampa, FL 33620

²Department of Biochemistry and Molecular Biophysics, and Biological Sciences, Columbia University, New York, 10032

³Department of Chemistry and Biology, Universidad del Norte, Barranquilla, Atlantico, Colombia

jasperbraun@mail.usf.edu

Complex genome rearrangements have been observed in certain organisms, such as the ciliated protist *Oxytricha trifallax*. These organisms are model organisms for studying DNA rearrangements. During the conjugation, the precursor genome contained in the germline micronucleus is reorganized to form the product genome of the somatic macronucleus. This process requires the elimination of large portions of the precursor genome and the reordering and inversion of many of the remaining precursor segments. To properly study the biological mechanism behind such massive rearrangements, good annotation of the two sequenced genomes is needed. We introduce terminology to describe the maps between precursor and product DNA sequences, and classify which of these maps are scrambled. An algorithm that computes the maps between the two genomes and determines those that are scrambled was developed as Scrambled DNA Rearrangement Annotation Pipeline (SDRAP) and tested on the precursor and product genomes of *O. trifallax*.