RNA-Guided DNA-Recombination Through Assembly Graphs

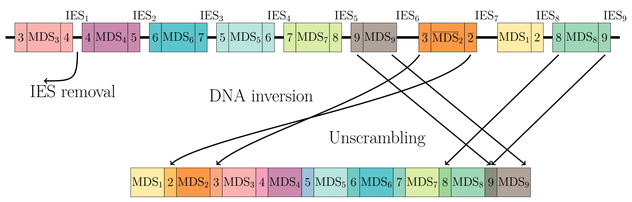
A. Angeleska, *University of Tampa*
E. Dolzhenko, N. Jonoska, M. Saito, *University of South Florida*
L. Landweber, *Princeton University*

Background

In a wide range of eukaryotes genome rearrangements have been observed on an evolutionary scale, as well as on a developmental scale, Even in cancer cells, dramatic DNA rearrangements are frequently observed in somatic cell lineages. We use gene rearrangements in ciliates *Oxytricha trifallax* as a model system to develop mathematical techniques and models to study DNA recombination events.

DNA Rearrangements in *O. trifallax*

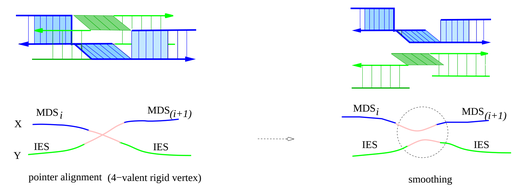
Ciliates undergo massive recombination during differentiation of an archival germline micronucleus into a somatic macronucleus capable of gene expression. The DNA processing events involve global deletion of 95-98% of the germline DNA, effectively eliminating *all* so-called "junk" DNA. These processes also eliminate hundreds of thousands of intervening DNA segments (internal eliminated sequences, IESs) that interrupt macronuclear genes which appear as several nonconsequtive segments (macronuclear destine segments, MDSs) in the micronucleus. Pointer-like sequences exist that are repeated at the end of each nth MDS and the beginning of each $(n+1)$st MDS in the micronucleus [4, 8].



Actin I micronuclear germline gene in *Oxytricha nova* (top) and the correctly assembled macronuclear gene (bottom)[8]. Each block represents an MDS, and each line between blocks is an IES. The numbers at the beginning and at the end of each segment represent the pointer sequences. The bars above MDS_2 and its pointers indicate that this block is inverted in the germline relative to the others.

Assembly Graphs

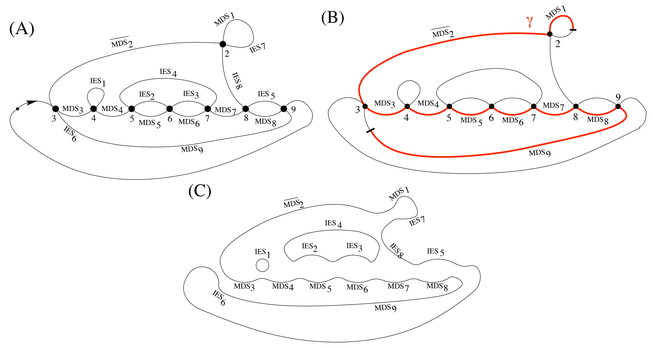
The proposed theoretical model utilizes spatial graphs as a physical representation of the DNA at the time of recombination, and smoothing of the vertices models the actual recombination.



(Left) top: schematic representation of two micronuclear DNA segments in DNA braiding position. Strands exchange pointer nucleotides through branch migration; bottom: the alignment represented as a 4-valent rigid vertex in a graph. **(Right)** top: two molecules after the recombination, with the MDSs joined on the same molecule; bottom: schematic representation of the finished recombination as smoothing of the vertex.

DNA Rearrangement with Assembly Graphs

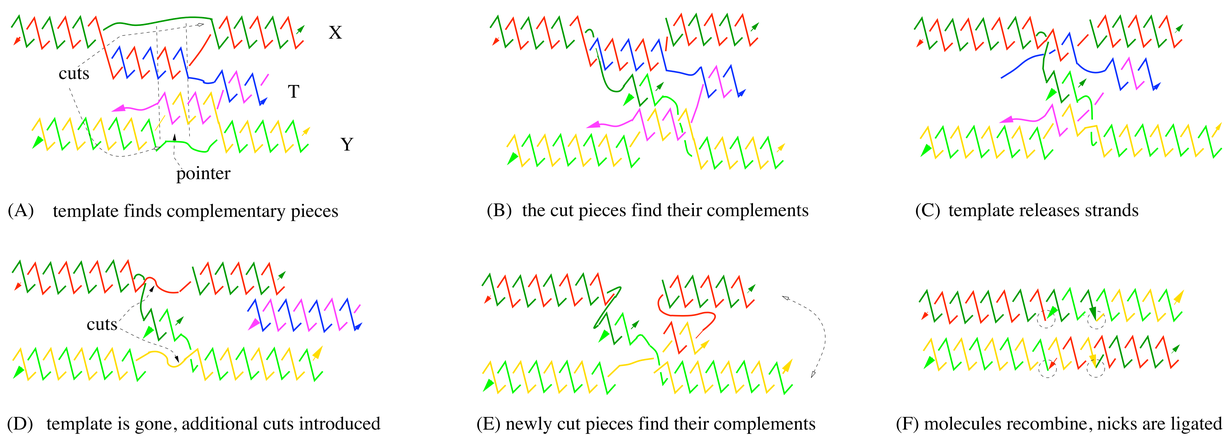
A micronuclear sequence is modeled with an *assembly graph* which is a finite connected graph with 4-valent rigid vertices. A macronuclear gene consisting of the ordered MDS segments are modeled with a *polygonal path*, an open path such that consecutive edges are neighbors with respect to the joint incident vertex.



(A) Graph structure of simultaneous recombination for *Actin I*, at the moment of recombination. Its edges are labeled with the micronuclear sequence of MDSs and IESs; (B) a polygonal path in the graph indicating MDSs for the macronuclear gene; (C) smoothing of the vertices guided by the polygonal path in (B) and the resulting molecules after recombination.

RNA-template Assembly

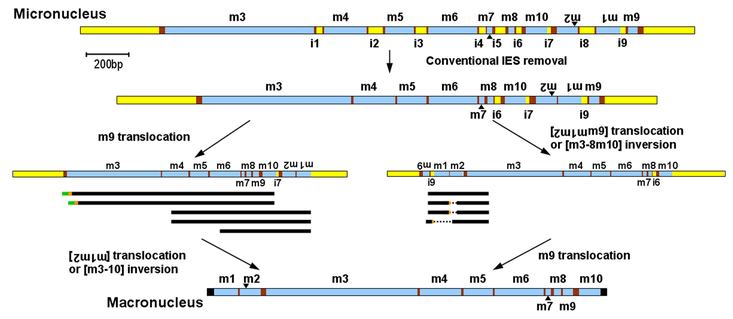
The process of DNA assembly, is guided by maternal RNA template sequences [1, 6].



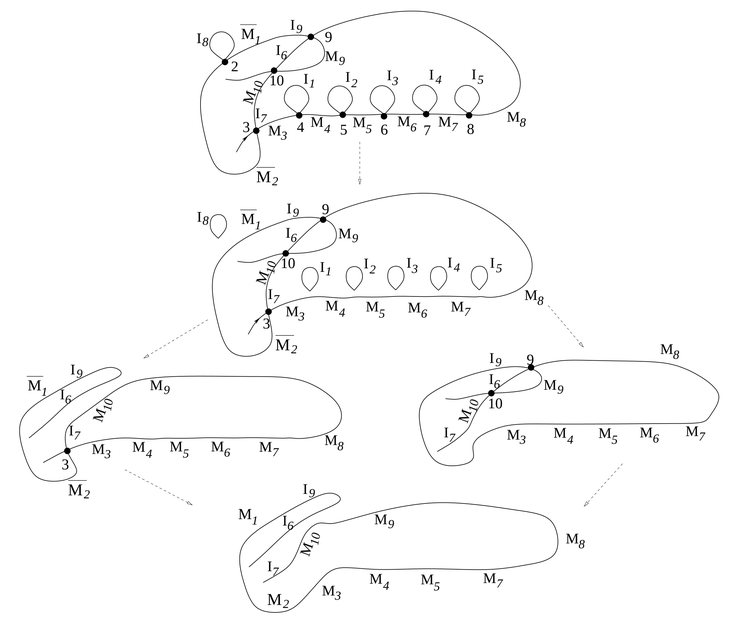
(A) Three molecules, two micronuclear segments X, Y and a dsRNA template T interact, permitting the template strands to find their corresponding complements in X and Y. Due to topological constraints we expect enzymatic cuts at this stage of the process [7]. (B) Through branch migration, the template displaces base-pairs between X and Y. A portion of X, containing the pointer sequence becomes single stranded and pairs with a portion of Y with the corresponding complementary sequence. (C) Branch migration begins, pairing between the template strands reinstates, releasing strands from X and Y. (D) Template base-pairs are restored, leaving the template unchanged. We postulate second pair of cuts at this time. (E) *DNA Braiding*: As X and Y dissociate from the template pairing between newly freed strands of X and Y develops. (F) Resulting recombined molecules (nicks are ligated).

Recombination Pathways

Recent data [5] confirmed a cascading process in descrambling. A model based on sets of formal symbols and circular strings was developed. With this model we can show that the observed set of intermediate sequences can be a result of two distinct recombination pathways of *Actin I* of *O. trifallax* [3].



Top: Two possible pathways to assemble the *S. lemnae Actin I* gene from its precursor form (m, MDS; i, IES) based on data from [5]. **Bottom:** Two theoretical assembly strategies for these pathways [3]. First S_1-smoothing is applied for the vertex set $S_1 = \{2, 4, 5, 6, 7, 8\}$, then S_2-smoothing is followed by S_3-smoothing or vice versa, where S_2 and S_3 are sets $S_2 = \{9, 10\}$ and $S_3 = \{3\}$



References

[1] A. Angeleska, N. Jonoska, M. Saito, L.F. Landweber, *J. of Theoretical Biology* 248(4) (2007) 706–720.

[2] A. Angeleska, N. Jonoska, M. Saito, *Discrete and Applied Math*, **157** (2009) 3020–3037.

[3] A. Angeleska, N. Jonoska, M. Saito, L.F. Landweber, *Recombination pathways in DNA rearrangements in ciliates,* in preparation.

[4] L.F. Landweber, T-C. Kuo, E.A. Curtis, *Proc. Nat. Acad. Sci. USA,* 97:7 (2000) 3298–3303.

[5] M. Mollenbeck, Y. Zhou, A.R.O. Cavalcanti, F. Jonsson, W.-J. Chang, S. Juranek, T.G. Doak, G. Rozenberg, H.J. Lipps, L.F. Landweber, (2008) PLoS ONE 3(6):e2330.

[6] M. Nowacki, V. Vijayan, Y. Zhou, T. Doak, E. Swart, L.F. Landweber, *Nature* (2008) 451:153–158.

[7] M. Nowacki, B.P. Higgins, G.M. Maquilan, E.C. Swart, T.G. Doak, L.F. Landweber. *Science* (2009) May 15;324(5929):935-8. Epub 2009 Apr 16.

[8] D.M. Prescott, A.F. Greslin, Dev Genet. 13:1 (1992) 66–74.