Micronuclear Sequences Associated with Assembly Graphs

Christeen Bisnath

September 20, 2011

Abstract

This paper investigates given an assembly graph, find the possible micronuclear sequences in terms of MDSs and IESs; which represent Hamiltonian polygonal paths. We will consider the orientations of the assembly graph and Hamiltonian polygonal path. To obtain a micronuclear sequence of this path, we will compare orientation of the Hamiltonian polygonal path with respect to the orientation of the assembly graph. We concentrate on two assembly graphs with two 4-valent vertices. These Hamiltonian polygonal paths influence the type of smoothings performed. There are two types of smoothings parallel and non-parallel smoothing. We consider a micronuclear sequence after one smoothing of a vertex.

1 Introduction

During the macronuclear development of ciliates *(Stylonychia or Oxytricha)*, DNA recombination occurs. Each nucleotide sequence of the macronucleus may be in non-consecutive order in the micronucleus as macronuclear destined sequence (MDSs) separated by non-coding DNA. During macronuclear development the non-coding fragments (internal eliminated sequence, IESs) that interrupt MDSs in the micronucleus are excised. However, the order of MDSs in the micronucleus may not be consecutive, and may be reversed relative to the others [1]. The rearrangement of a micronuclear sequence can be seen as MDSs and IESs. The assembly graphs represent these micronuclear genes. An assembly graph is a finite connected graph where all vertices are rigid vertices of valency 1 or 4 [1]. We will focus on assembly graphs that have Hamiltonian polygonal paths that represent a single gene combination in the micronucleus, such graphs are called *realizable graphs*. A polygonal path is a path in which every two consecutive edges are neighbors with respect to their common vertex or in other words, makes a "90 degree" turn at every rigid vertex [1]. A Hamiltonian polygonal path is a path in which each 4-valent is visited exactly once, makes a "90 degree turn" at each rigid vertex and no edges are repeated. A transverse path is a path in which every two consecutive edges are not neighbors. The given assembly graph (transversal) corresponds to the micronuclear DNA segment before recombination. A transverse graph (transversal) is where a single path contains every edge is visited and none is repeated, which are called simple assembly graphs [3]. We are interested in what types of rearrangements and how many distinct genes can be encoded by an assembly graph [4].

2 Forming Micronuclear Sequences

2.1 Notation for Hamiltonian Polygonal Paths and Micronuclear Sequences

Given a transversal with possible Hamiltonian polygonal paths (HPPs), we will write possible DNA recombinations of micronuclear sequences. In order to understand Hamiltonian polygonal paths

1

and micronuclear sequences, the following will be some notation to help describe them.

- n represent 4-valent vertices

- $e_0, \ldots, e_{2n}$ represent edges, are enumerated by the way they are encountered by the transverse path, this fixes an orientation of the edge

- $v_0, \ldots, v_{n+1}$ represent vertices of assembly graph

- $\Gamma = \gamma_1, \ldots, \gamma_{n+1}$ where Γ is one HPP and γ_1, γ_{n+1} are ending edges and $\gamma_2, \ldots, \gamma_n$ are full edges of G and Γ, let G assembly graph be n 4-valent vertices

- $P = \gamma_2, \ldots, \gamma_n$

- $I_0, \ldots, I_{n+1}$ represent IES sequences, for n is number of vertices in assembly graph

- $M_1, \ldots, M_{n+1}$ represent MDS sequences, for n is number of vertices in assembly graph

- $\overline{M_1}, \ldots, \overline{M_{n+1}}$ represent inverted MDS sequences, for n is number of vertices in assembly graph

- - - - - - - - - - - dashed lines represent the Hamiltonian polygonal path

- _____ solid line represent the transversal path

- A= incoming edges, B= outgoing edges; of transversal path orientation
  We indicate with A an incoming edge incident to a vertex and with B outgoing edge.

Now, with a given assembly graph and Hamiltonian polygonal path without ending edges (P), consider the first and last ending edges of the path, (γ_1, γ_{n+1}) and vertices of the path. For P= $\gamma_2, \ldots, \gamma_n$, which are the complete edges from the assembly graph. Also, HPP= $w_0 \, \gamma_1, \ldots, w_{n-1} \gamma_{n+1}$; the order of encountering these edges defines an orientation of the HPP. Label the incoming and outgoing edges that are neighbors of γ_1 incident to v_{i_0} with 1A and 1B respectively. Then, label the incoming and outgoing edges that are neighbors of γ_{n+1} incident to v_{i_n} with 2A and 2B respectively, for some i. Each choice of labels 1A, 1B or 2A, 2B indicate a possible choice of "start" or "end" of an MDS sequence. To denote a specific HPP with chosen ends path, use $P(\gamma_1, \gamma_{n+1})$. Each γ_1 and γ_{n+1} will be either one of the labels of ending edges and P is composed of complete edges. This path will take upon its own orientation; separate from the orientation of the assembly graph. To obtain the micronuclear sequence, compare the orientation of the HPP to the original path orientation of the assembly graph. These orientations will indicate the MDSs' orientation. Let, $\lambda(e_i)$ be the label for e_i for some i, then the micronuclear sequence corresponds to $\lambda(e_o), \ldots, \lambda(e_{2n})$. All other edges not visited by HPP are labeled with IESs. Note, an edge may be both 1A and 2B or 1B and 2A (See Figure 7). Such edge is not in the list of HPP edges, $\gamma_1, \ldots, \gamma_{n+1}$. Hence, it also has a label I_k for an IES, for some k, so one of the edges may have up to three labels 1A,

I_k, 2B for example, see Figure 2. In the micronuclear sequence, if the orientation of γ_i obtained from the transversal is opposite, the orientation γ_i obtains from HPP, label γ_i with $\overline{M_i}$ otherwise, its label remains M_i. We assume that the assembly graphs are open ended, so we will always start and end with IESs in the micronuclear sequences. But, to find all possible micronuclear sequences, we will find all possible combinations of orientations of each HPP with orientations of assembly graph. So, each HPP has four distinct combinations of orientations, with respect to orientation of the HPP and assembly graph. The four distinct orientations are Γ, Γ^R, Γ^- and Γ^{-R}. Γ is an HPP, $P(\gamma_1, \gamma_{n+1})$. Γ^R is the opposite orientation of the assembly graph and same HPP orientation. Γ^- is the same orientation as the assembly graph and opposite orientation of HPP. Γ^{-R} is both opposite orientations of assembly graph and HPP. To obtain the micronuclear sequences of the other 3 distinct orientations, we go through a similar process of the HPPs' of the following figures, once the orientations of the assembly graph and HPP are fixed. The following HPPs illustrations are Γ, with its corresponding micronuclear sequence.

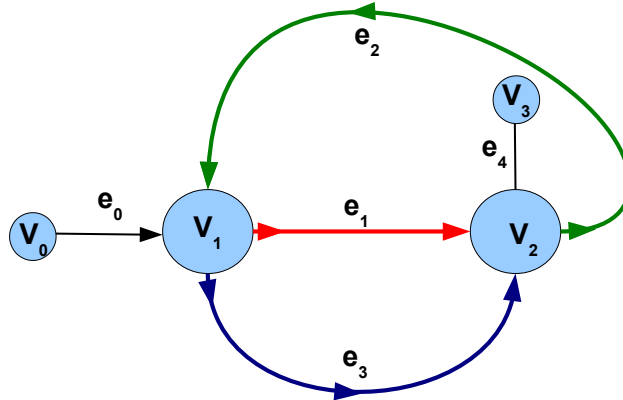## 2.2 Possible Micronuclear Sequences Represented by Assembly Graph 1212

Figure 1: Assembly graph 1212 consists of the two 4-valent vertices and five edges and transverse path starts with $v_0, \ldots, v_3$. For example, the graph above starts with an edge e_0 that first visits v_1, v_2, v_1, v_2, giving us assembly graph 1212.

$1)e_2(1A, 2A)$
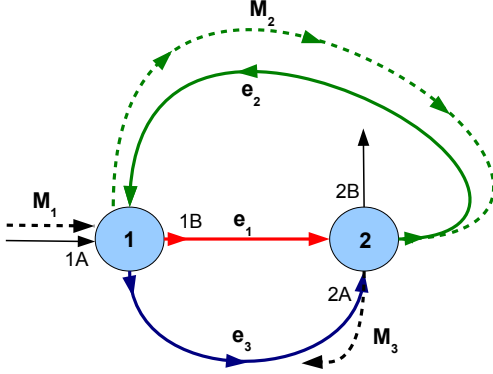$I_0 M_1 I_1 \overline{M_2} I_2 \overline{M_3} I_3$

$2)e_2(1A, 2B)$
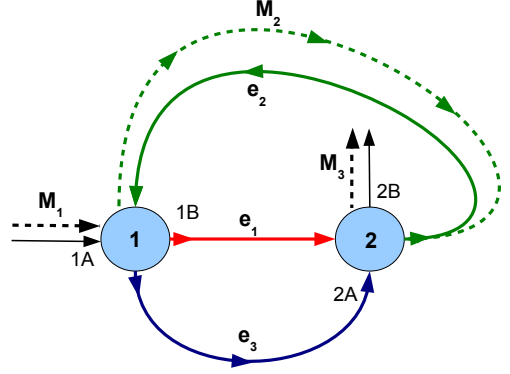$I_0 M_1 I_1 \overline{M_2} I_2 M_3 I_3$

Figure 2

Figure 3

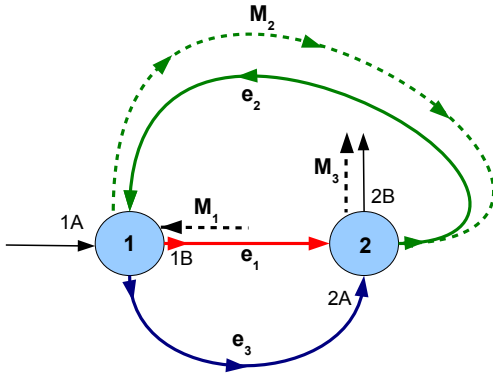$3)e_2(1B, 2B)$
$I_0 \overline{M_1} I_1 \overline{M_2} I_2 M_3 I_3$

$4)e_2(1B, 2A)$
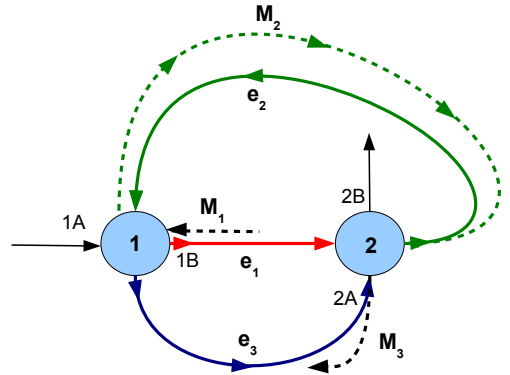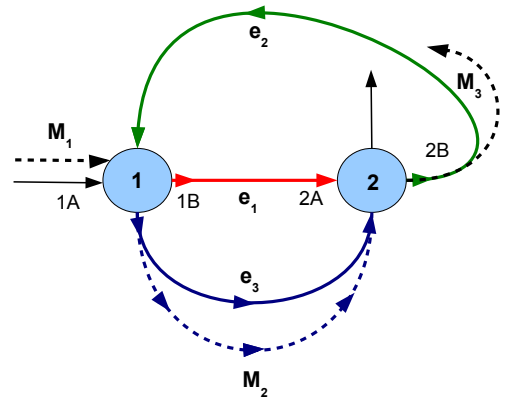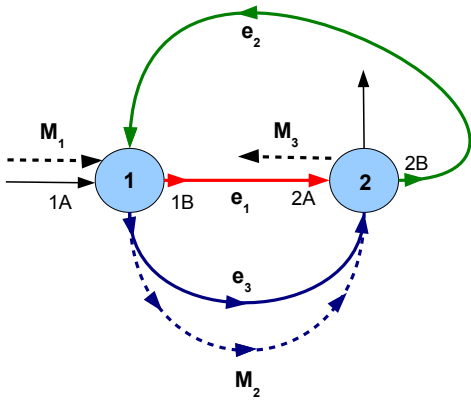$I_0 \overline{M_1} I_1 \overline{M_2} I_2 \overline{M_3} I_3$

Figure 4

Figure 5

The figures above are HPPs (Γ) grouped consisting of the path (e_2).

Figure 2 has two vertices so there is 1 full edge in the HPP (e_2), with ending edges $(1A, 2A)$. Full edge, in the sense that it is a complete edge of the assembly graph and Γ. So, the first line of notation above Figure 2 denotes HPP as $e_2(1A, 2A)$. The second line of notation above the assembly graph represents the micronuclear sequence of the Hamiltonian polygonal path. To obtain the micronuclear sequence, read the labels of the transversal path of the given graph. For example, we always start with an IES (I_0) following $1A$, which is the first MDS (M_1) and continues along e_1, which is I_1. Now, moving along e_2 ($\overline{M_2}$) we continue to I_2 which is part of e_3. Then, with $\overline{M_3}$ $(2A)$, however, we must complete the assembly graph and micronuclear sequence with an IES (I_3).

4

Notice, the MDS sequences in the HPP are inverted, because their orientations are the opposite with respect to the transversal. Obtaining the micronuclear sequences for Γ^R, Γ^- and Γ^{-R} is the same process as obtaining Γ, as long as the combinations of orientations for each are distinct, with respect to the assembly graph and HPP.

$$5)e_3(1A, 2A)$$
$$I_0 M_1 I_1 \overline{M_3} I_2 M_2 I_3$$

$$6)e_3(1A, 2B)$$
$$I_0 M_1 I_1 M_3 I_2 M_2 I_3$$

Figure 6

Figure 7

$$7)e_3(1B, 2A)$$
$$I_0 \overline{M_1} I_1 \overline{M_3} I_2 M_2 I_3$$

$$8)e_3(1B, 2B)$$
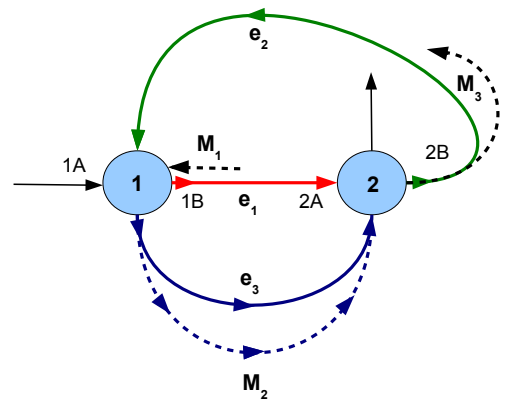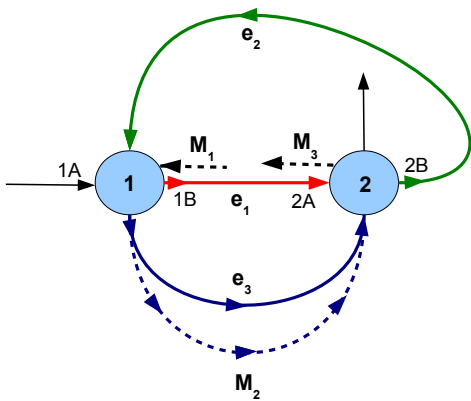$$I_0 \overline{M_1} I_1 M_3 I_2 M_2 I_3$$

Figure 8

Figure 9

The figures above are HPPs (Γ) grouped consisting of the path (e_3).

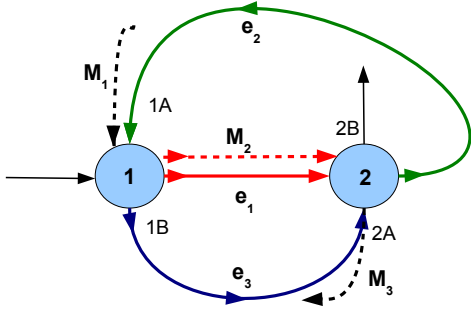9)$e_1(1A, 2A)$
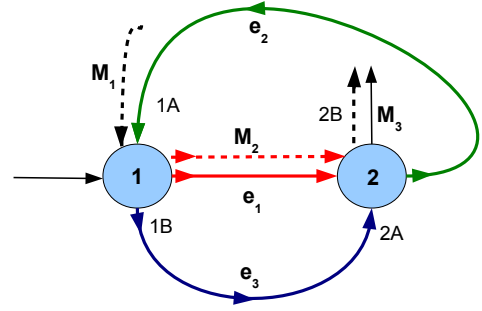$I_0M_2I_1M_1I_2\overline{M_3}I_3$

10)$e_1(1A, 2B)$
$I_0M_2I_1M_1I_2M_3I_3$

Figure 10

Figure 11

11)$e_1(1B, 2A)$
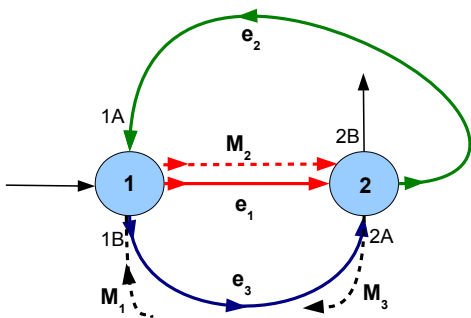$I_0M_2I_1\overline{M}_1I_2\overline{M_3}I_3$

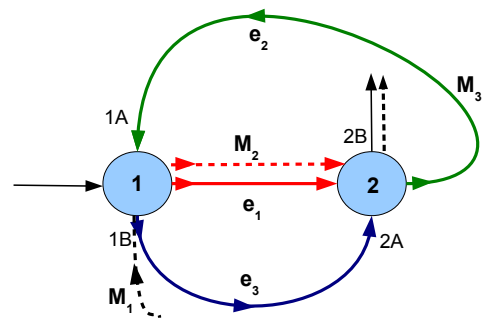12)$e_1(1B, 2B)$
$I_0M_2I_1\overline{M}_1I_2M_3$

Figure 12

Figure 13

The figures above are HPPs (Γ)grouped consisting of the path (e_1).

HPP	Micronuclear Sequence	Inverted Micronuclear Sequence
1)$e_2(1A, 2A)$	$I_0 M_1 I_1 \overline{M_2} I_2 \overline{M_3} I_3$	$I_0 M_3 I_1 M_2 I_2 \overline{M_1} I_3$
2)$e_2(1A, 2B)$	$I_0 M_1 I_1 \overline{M_2} I_2 M_3 I_3$	$I_0 \overline{M_3} I_1 M_2 I_2 \overline{M_1} I_3$
3)$e_2(1B, 2B)$	$I_0 \overline{M_1} I_1 \overline{M_2} I_2 M_3 I_3$	$I_0 \overline{M_3} I_1 M_2 I_2 M_1 I_3$
4)$e_2(1B, 2A)$	$I_0 \overline{M_1} I_1 \overline{M_2} I_2 \overline{M_3} I_3$	$I_0 M_3 I_1 M_2 I_2 M_1 I_3$
5)$e_3(1A, 2A)$	$I_0 M_1 I_1 \overline{M_3} I_2 M_2 I_3$	$I_0 \overline{M_2} I_1 M_3 I_2 \overline{M_1} I_3$
6)$e_3(1A, 2B)$	$I_0 M_1 I_1 M_3 I_2 M_2 I_3$	$I_0 \overline{M_2} I_1 \overline{M_3} I_2 \overline{M_1} I_3$
7)$e_3(1B, 2A)$	$I_0 \overline{M_1} I_1 \overline{M_3} I_2 M_2 I_3$	$I_0 \overline{M_2} I_1 M_3 I_2 M_1 I_3$
8)$e_3(1B, 2B)$	$I_0 \overline{M_1} I_1 M_3 I_2 M_2 I_3$	$I_0 \overline{M_2} I_1 \overline{M_3} I_2 M_1 I_3$
9)$e_1(1A, 2A)$	$I_0 M_2 I_1 M_1 I_2 \overline{M_3} I_3$	$I_0 M_3 I_1 \overline{M_1} I_2 \overline{M_2} I_3$
10)$e_1(1A, 2B)$	$I_0 M_2 I_1 M_1 I_2 M_3 I_3$	$I_0 \overline{M_3} I_1 \overline{M_1} I_2 \overline{M_2} I_3$
11)$e_1(1B\ \ 2A)$	$I_0 M_2 I_1 \overline{M_1} I_2 \overline{M_3} I_3$	$I_0 M_3 I_1 M_1 I_2 \overline{M_2} I_3$
12)$e_1(1B\ \ 2B)$	$I_0 M_2 I_1 \overline{M_1} I_2 M_3 I_3$	$I_0 \overline{M_3} I_1 M_1 I_2 \overline{M_2} I_3$

The first column in the table above are the possible HPPs (Γ) of assembly graph 1212. The second column represent the micronuclear sequences obtained from the HPPs. The third column is Γ^R. Some sequences in Γ^- and Γ^{-R} corresponded to Γ, Γ^R, which leaves us with twenty four distinct labellings of the assembly graph 1212. Notice, that Γ^R is complementary to the original, Γ and therefore represent the same DNA sequence [2]. Graphically, the difference between Γ^R and Γ is the orientation of the assembly graph, which inverts the sequences when obtained.

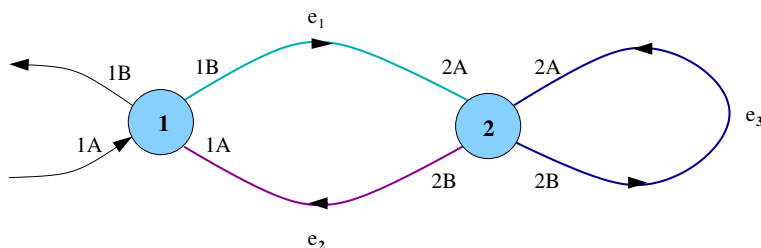## 2.3   Possible Micronuclear Sequences Represented by Assembly Graph 1221

Figure 2: The assembly graph above is 1221, which is formed similarly to the previous assembly graph.
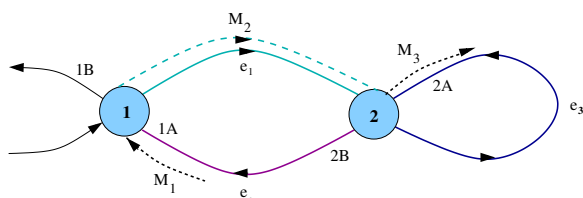
1)$e_1(1A, 2A)$
$I_0 M_2 I_1 \overline{M_3} I_2 M_1 I_3$

Figure 2
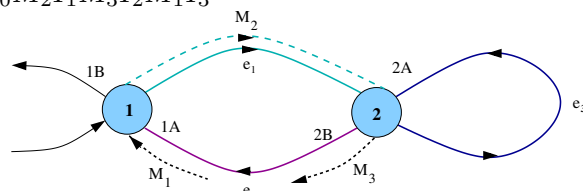
2)$e_1(1A, 2B)$
$I_0 M_2 I_1 M_3 I_2 M_1 I_3$

Figure 3

3)$e_1(1B, 2A)$
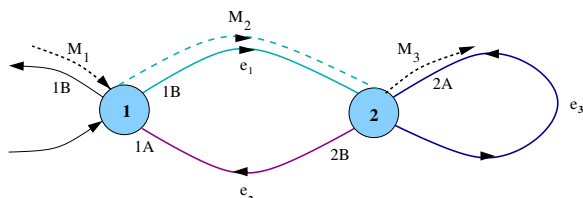$I_0 M_2 I_1 \overline{M_3} I_2 \overline{M_1} I_3$

Figure 4

4)$e_1(1B, 2B)$
$I_0 M_2 I_1 M_3 I_2 \overline{M_1} I_3$
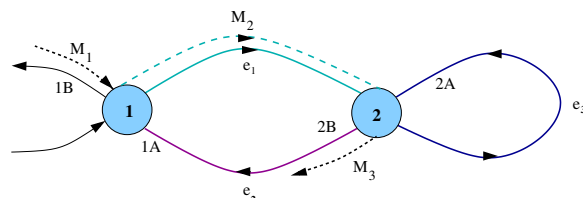

Figure 5

The figures above are HPPs (Γ) grouped consisting of the path (e_1). The corresponding micronuclear sequences are obtained in a similar fashion as the previous micronuclear sequences of previous assembly graph 1212.
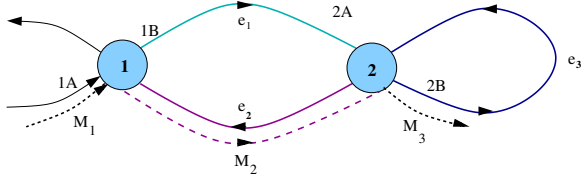
5)$e_2(1A, 2B)$
$I_0 M_1 I_1 M_3 I_2 \overline{M_2} I_3$

Figure 6

6)$e_2(1A, 2A)$
$I_0 M_1 I_1 \overline{M_3} I_2 \overline{M_2} I_3$

Figure 7

7)$e_2(1B, 2A)$
$I_0 \overline{M_1} I_1 \overline{M_3} I_2 \overline{M_2} I_3$

Figure 8

8)$e_2(1B, 2B)$
$I_0 \overline{M_1} I_1 M_3 I_2 \overline{M_2} I_3$

Figure 9

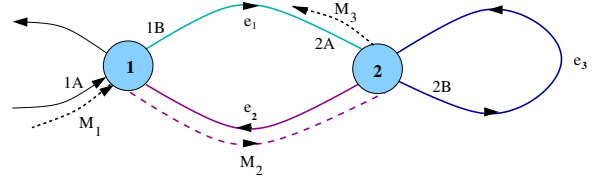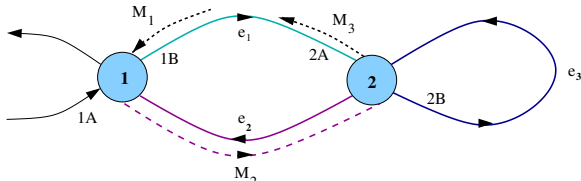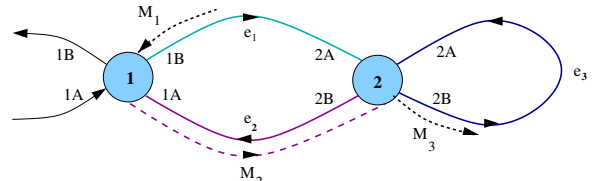The figures above are HPPs (Γ) grouped consisting of the path (e_2). The corresponding micronuclear sequences are obtained in a similar fashion as the previous micronuclear sequences of previous assembly graph 1212.

HPP	Micronuclear Sequence	Inverted Micronuclear Sequence
1)$e_1(1A, 2A)$	$I_0 M_2 I_1 \overline{M_3} I_2 M_1 I_3$	$I_0 \overline{M_1} I_1 M_3 I_2 \overline{M_2} I_3$
2)$e_1(1A, 2B)$	$I_0 M_2 I_1 M_3 I_2 M_1 I_3$	$I_0 \overline{M_1} I_1 \overline{M_3} I_2 \overline{M_2} I_3$
3)$e_1(1B, 2A)$	$I_0 M_2 I_1 \overline{M_3} I_2 \overline{M_1}$	$M_1 I_0 M_3 I_1 \overline{M_2} I_2$
4)$e_1(1B, 2B)$	$I_0 M_2 I_1 M_3 I_2 \overline{M_1}$	$M_1 I_0 \overline{M_3} I_1 \overline{M_2} I_2$
5)$e_2(1A, 2B)$	$I_0 M_1 I_1 M_3 I_2 \overline{M_2} I_3$	$I_0 M_2 I_1 \overline{M_3} I_2 \overline{M_1} I_3$
6)$e_2(1A, 2A)$	$I_0 M_1 I_1 \overline{M_3} I_2 \overline{M_2} I_3$	$I_0 M_2 I_1 M_3 I_2 \overline{M_1} I_3$
7)$e_2(1B, 2A)$	$I_0 \overline{M_1} I_1 \overline{M_3} I_2 \overline{M_2} I_3$	$I_0 M_2 I_1 M_3 I_2 M_1 I_3$
8)$e_2(1B, 2B)$	$I_0 \overline{M_1} I_1 M_3 I_2 \overline{M_2} I_3$	$I_0 M_2 I_1 \overline{M_3} I_2 M_1 I_3$

The first column in the table above are the possible HPPs (Γ) of assembly graph 1221. The second column represent the micronuclear sequences obtained from the HPPs. The third column is Γ^R. Notice, the sequences of Γ^R are complementary to the DNA sequences of Γ. Which leaves eight distinct micronuclear sequences of this fixed orientation. For example:

The Γ^R sequence of line 8 is the sequence in line 1

The Γ^R sequence of line 7 is the sequence in line 2

The Γ^R sequence of line 5 is the sequence in line 3

The Γ^R sequence of line 6 is the sequence in line 4

But, we have to find all possible micronuclear sequences for assembly graph 1221, which also considers Γ^{-R} and Γ^-. The other eight distinct micronuclear sequences come from Γ^{-R} and Γ^- as follows,

HPP Orientation	Micronuclear Sequence
Γ^-	$I_0 \overline{M_2} I_1 M_1 I_2 \overline{M_3} I_3$
	$I_0 \overline{M_2} I_1 \overline{M_1} I_2 \overline{M_3} I_3$
	$I_0 \overline{M_2} I_1 M_1 I_2 M_3 I_3$
	$I_0 \overline{M_2} I_1 \overline{M_1} I_2 M_3 I_3$
Γ^{-R}	$I_0 M_3 I_1 \overline{M_1} I_2 M_2 I_3$
	$I_0 M_3 I_1 M_1 I_2 M_2 I_3$
	$I_0 \overline{M_3} I_1 \overline{M_1} I_2 M_2 I_3$
	$I_0 \overline{M_3} I_1 M_1 I_2 M_2 I_3$

with a total of sixteen distinct micronuclear sequences of assembly graph 1221.

3 Smoothing of Assembly Graphs

Smoothings of assembly graphs represent alignment of pointers (MDS pointers) in the micronuclear sequences and the rearrangement. Each vertex, there are edges incident to the vertex labeled with MDSs and IESs that cross each vertex. When smoothed at the vertex (removing the vertex), the pointers align (See Figure 10). During this process, MDSs align and the in the micronuclear sequence, recombine MDSs and IESs. There are two types of smoothing, N (non-parallel) smoothing and P (parallel) smoothing (See Figure 11). If the pointer sequences are aligned in parallel, then the smoothing at the vertex follows the predetermined direction of the assembly graph (parallel smoothing). If the pointer alignment is anti parallel, then the smoothing of the vertex changes the path of orientation of the original graph [1]. Smoothing of each vertex in an assembly graph, is determined by the HPPs of the assembly graph. The type of smoothing is predetermined by the given HPP, in such a way the HPP stays connected. The writing of the "new" micronuclear sequences after the smoothing, is similar to writing of the micronuclear sequence before the smoothing. Use similar notation from Sec.3 to write "new" micronuclear sequences. The difference from the original transversal "old" micronuclear sequence to the new assembly graph "new" micronuclear sequence (not necessary a transversal) is recombined MDSs, since that's what smoothing does to the sequences. Notice, the appearance during P smoothings, the assembly graph is broken up into components, that is why there are bracket notations. The brackets show a separation in MDSs within the assembly graph.
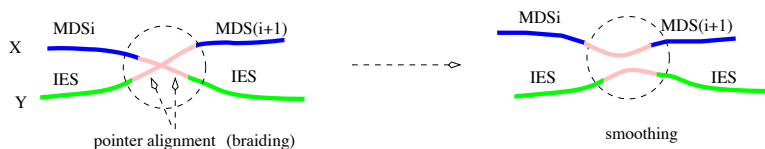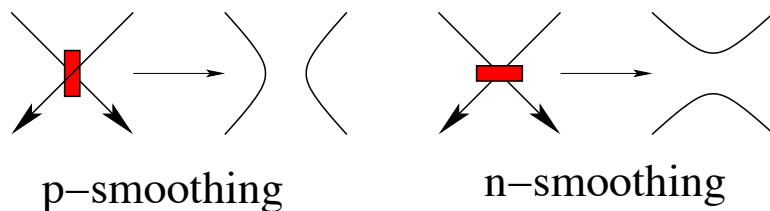

Figure 10

p–smoothing   n–smoothing

Figure 11

3.1 Smoothing of Assembly Graph 1212

The following figures represent smoothings of HPPs on assembly graph 1212. Where corresponding micronuclear sequences of the HPPs is on left side of arrow. The smoothings now correspond to a new micronuclear sequence, which is represented on the right side of the arrow. As stated previously, the HPPs predetermines the type of smoothing of the HPPs on the assembly graph, which is one reason why the sequences are grouped in this manner. The other is with respect to which vertex is removed. Notice, the merging of MDSs after vertex removal.

N smoothing at vertex 1

1) $I_0 M_1 I_1 \overline{M_2} I_2 \overline{M_3} I_3 \longrightarrow I_0 M_{1,2} I_1 \overline{M_3} I_2$
2) $I_0 M_1 I_1 \overline{M_2} I_2 M_3 I_3 \longrightarrow I_0 M_{1,2} I_1 M_3 I_2$
7) $I_0 \overline{M_1} I_1 \overline{M_3} I_2 M_2 I_3 \longrightarrow I_0 M_{3,1,2} I_1$
8) $I_0 \overline{M_1} I_1 M_3 I_2 M_2 I_3 \longrightarrow I_0 \overline{M_3} I_1 M_{1,2} I_2$
11) $I_0 M_2 I_1 \overline{M_1} I_2 \overline{M_3} I_3 \longrightarrow I_0 \overline{M}_{2,1,3} I_1$
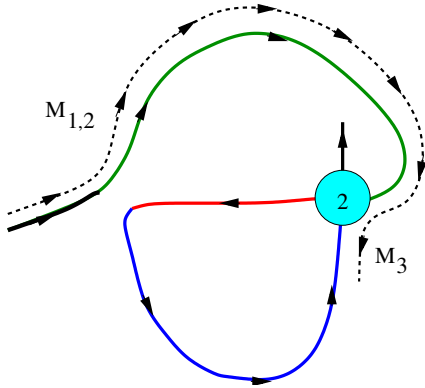12) $I_0 M_2 I_1 \overline{M_1} I_2 M_3 \longrightarrow I_0 \overline{M}_{2,1} I_1 M_3 I_2$

P smoothing at vertex 1

3) $I_0 \overline{M_1} I_1 \overline{M_2} I_2 M_3 \longrightarrow I_0 M_3, [I_1 \overline{M_{2,1}} I_2]$
4) $I_0 \overline{M_1} I_1 \overline{M_2} I_2 \overline{M_3} I_3 \longrightarrow I_0 \overline{M_3} I_1, [\overline{M_{2,1}} I_2]$
5) $I_0 M_1 I_1 \overline{M_3} I_2 M_2 I_3 \longrightarrow I_0 M_{1,2} I_1 [\overline{M_3} I_2]$
6) $I_0 M_1 I_1 M_3 I_2 M_2 I_3 \longrightarrow I_0 M_{1,2} I_1 [M_3 I_2]$
9) $I_0 M_2 I_1 M_1 I_2 \overline{M_3} I_3 \longrightarrow I_0 \overline{M_3} I_1 [M_{1,2} I_2]$
10) $I_0 M_2 I_1 M_1 I_2 M_3 I_3 \longrightarrow I_0 M_3 [I_1 M_{1,2} I_2]$

Figure 3: The following smoothing corresponds to the first micronuclear sequence of assembly graph 1212.
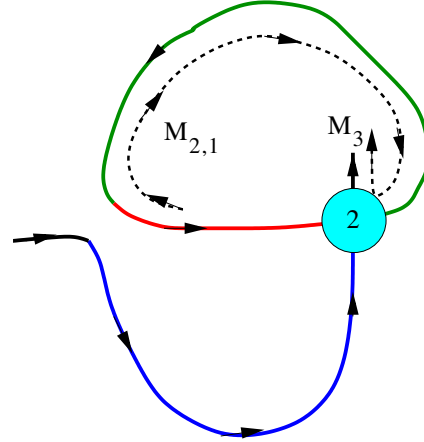

Figure 4: The following smoothing corresponds to the third micronuclear sequence of assembly graph 1212.

As seen in Figure 3, represents the assembly graph for first micronuclear sequence of assembly 1212, when vertex one is removed. We can see in the micronuclear sequence before the vertex removal, had separate MDSs. After, the vertex removal, there were merging of MDSs, M_1 and M_2 to $M_{1,2}$. To obtain the micronuclear sequence, read the transversal path with respect to the smoothed HPP, considering the fixed orientations of both.

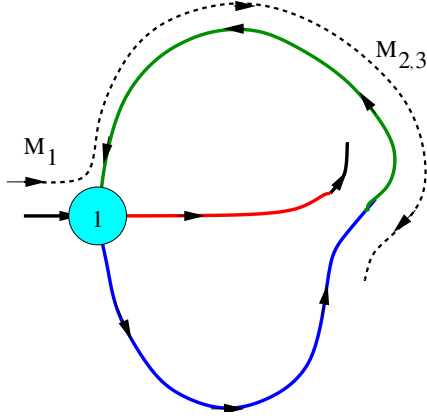|						
1)	$I_0 M_1 I_1 \overline{M_2} I_2 \overline{M_3} I_3$	$\longrightarrow$	$I_0 M_1 I_1 [\overline{M_{2,3}} I_2]$			
4)	$I_0 \overline{M_1} I_1 \overline{M_2} I_2 \overline{M_3} I_3$	$\longrightarrow$	$I_0 \overline{M_1} I_1 [\overline{M_{2,3}} I_2]$			
6)	$I_0 M_1 I_1 M_3 I_2 M_2 I_3$	$\longrightarrow$	$I_0 M_1 I_1 [M_{3,2} I_2]$			
8)	$I_0 \overline{M_1} I_1 M_3 I_2 M_2 I_3$	$\longrightarrow$	$I_0 \overline{M_1} I_1 [M_{3,2} I_2]$			
10)	$I_0 M_2 I_1 M_1 I_2 M_3 I_3$	$\longrightarrow$	$I_0 M_{2,3} [I_1 M_1 I_2]$			
12)	$I_0 M_2 I_1 \overline{M_1} I_2 M_3 I_3$	$\longrightarrow$	$I_0 M_{2,3} [I_1 \overline{M_1} I_2]$			

2)	$I_0 M_1 I_1 \overline{M_2} I_2 M_3 I_3$	$\longrightarrow$	$I_0 M_1 I_1 M_{2,3} I_2$
3)	$I_0 \overline{M_1} I_1 \overline{M_2} I_2 M_3 I_3$	$\longrightarrow$	$I_0 \overline{M_1} I_1 M_{2,3} I_2$
5)	$I_0 M_1 I_1 \overline{M_3} I_2 M_2 I_3$	$\longrightarrow$	$I_0 M_1 I_1 \overline{M_{3,2}} I_2$
7)	$I_0 \overline{M_1} I_1 \overline{M_3} I_2 M_2 I_3$	$\longrightarrow$	$I_0 \overline{M_{1,3,2}} I_1$
9)	$I_0 M_2 I_1 M_1 I_2 \overline{M_3} I_3$	$\longrightarrow$	$I_0 M_{2,3} I_1 \overline{M_1} I_2$
11)	$I_0 M_2 I_1 \overline{M_1} I_2 \overline{M_3} I_3$	$\longrightarrow$	$I_0 M_{2,3,1} I_1$

Figure 5: The following smoothing corresponds to the first micronuclear sequence of assembly graph 1212.
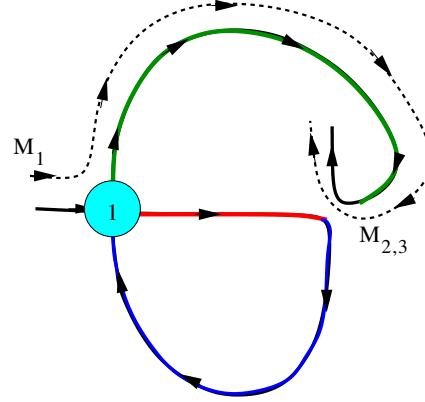

Figure 6: The following smoothing corresponds to the second micronuclear sequence of assembly graph 1212.

As seen in Figure 5, represents the assembly graph for first micronuclear sequence of assembly 1212, when vertex two is removed. We can see in the micronuclear sequence before the vertex removal, had separate MDSs. After, the vertex removal, there were merging of MDSs, $\overline{M_2}$ and $\overline{M_3}$ to $\overline{M_{2,3}}$. Notice, there are 2 components of the transversal. To obtain the micronuclear sequence, read the transversal path with respect to the smoothed HPP, considering the fixed orientations of both. The brackets are used to indicate the beginning of the second component of the transversal, with respect to the "new" micronuclear sequence.

3.2 Smoothing of Assembly Graph 1221

The following figures represent smoothings of HPPs on assembly graph 1221. Where corresponding micronuclear sequences of the HPPs is on left side of arrow. The smoothings now correspond to a new micronuclear sequence, which is represented on the right side of the arrow. As stated previously, the HPPs predetermines the type of smoothing of the HPPs on the assembly graph, which is one reason why the sequences are grouped in this manner. The other is with respect to which vertex is removed. Notice, the merging of MDSs after vertex removal. Obtaining the "new" micronuclear sequences is similar to smoothings of assembly graph 1212.

P smoothing at vertex 1

$$
\begin{array}{lll}
1) & I_0 M_2 I_1 \overline{M_3} I_2 M_1 I_3 & \longrightarrow & I_0 M_{1,2} I_1 \overline{M_3} I_2 \\
2) & I_0 M_2 I_1 M_3 I_2 M_1 I_3 & \longrightarrow & I_0 M_{1,2} I_1 M_3 I_2 \\
7) & I_0 \overline{M_1} I_1 \overline{M_3} I_2 \overline{M_2} I_3 & \longrightarrow & I_0 \overline{M_{2,1,3}} I_1 \\
8) & I_0 \overline{M_1} I_1 M_3 I_2 \overline{M_2} I_3 & \longrightarrow & I_0 \overline{M_{2,1}} I_1 M_3 I_2
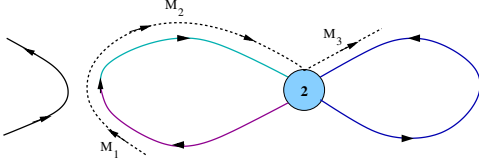\end{array}
$$

Figure 7: The following smoothing corresponds to the first micronuclear sequence of assembly graph 1221.

N smoothing at vertex 1

$$
\begin{array}{lll}
3) & I_0 M_2 I_1 \overline{M_3} I_2 \overline{M_1} I_3 & \longrightarrow & I_0 M_3 I_1 \overline{M_{2,1}} I_2 \\
4) & I_0 M_2 I_1 M_3 I_2 \overline{M_1} I_0 & \longrightarrow & I_0 \overline{M_3} I_1 \overline{M_{2,1}} I_2 \\
5) & I_0 M_1 I_1 M_3 I_2 \overline{M_2} I_3 & \longrightarrow & I_0 M_{1,2} I_1 \overline{M_3} I_2 \\
6) & I_0 M_1 I_1 \overline{M_3} I_2 \overline{M_2} I_3 & \longrightarrow & I_0 M_{1,2} I_1 M_3 I_2
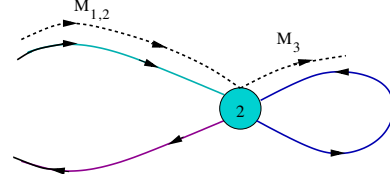\end{array}
$$

Figure 8: The following smoothing corresponds to the third micronuclear sequence of assembly graph 1221. Removing vertex 1 allowed $\overline{M_1}$ and M_2 to align forming $\overline{M_2, 1}$.

P smoothing at vertex 2

$$
\begin{array}{lll}
2) & I_0 M_2 I_1 M_3 I_2 M_1 I_3 & \longrightarrow & I_0 M_{2,3} I_1 M_1 I_2 \\
4) & I_0 M_2 I_1 M_3 I_2 \overline{M_1} I_3 & \longrightarrow & I_0 M_{2,3} I_1 \overline{M_1} I_2 \\
6) & I_0 M_1 I_1 \overline{M_3} I_2 \overline{M_2} I_3 & \longrightarrow & I_0 M_1 I_1 \overline{M_{3,2}} I_2 \\
7) & I_0 \overline{M_1} I_1 \overline{M_3} I_2 \overline{M_2} I_3 & \longrightarrow & I_0 \overline{M_1} I_1 \overline{M_{3,2}} I_2
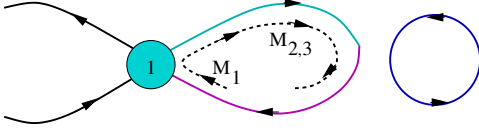\end{array}
$$

Figure 9: The following smoothing corresponds to the second micronuclear sequence of assembly graph 1221.

N smoothing at vertex 2

$$
\begin{array}{lll}
1) & I_0 M_2 I_1 \overline{M_3} I_2 M_1 I_3 & \longrightarrow & I_0 M_{2,3} I_1 M_1 I_2 \\
3) & I_0 M_2 I_1 \overline{M_3} I_2 \overline{M_1} I_3 & \longrightarrow & I_0 M_{2,3} I_1 \overline{M_1} I_2 \\
5) & I_0 M_1 I_1 M_3 I_2 \overline{M_2} I_3 & \longrightarrow & I_0 M_1 I_1 \overline{M_{3,2}} I_2 \\
8) & I_0 \overline{M_1} I_1 M_3 I_2 \overline{M_2} I_3 & \longrightarrow & I_0 \overline{M_1} I_1 \overline{M_{3,2}} I_2
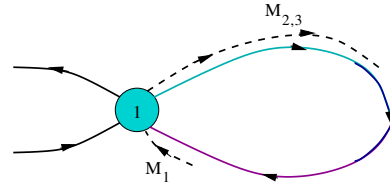\end{array}
$$

Figure 10: The following smoothing corresponds to the second micronuclear sequence of assembly graph 1221.

4 Conclusion

The forming of micronuclear sequences through HPP can influence the resulting sequences. Such as, the smoothing of vertices, which behave in a manner to preserve the HPP. But, at the same time eliminate some IESs and align MDSs. The previous figures were for two vertices, as the vertices increases, the HPP gets more complicated. The micronuclear sequences are longer, due to the fact more edges are used to form the HPP. The smoothings are more abstract; since the assembly graphs are larger and more smoothings since there are more vertices. There must be four distinct micronuclear sequences for a given path $P(P(1A, 2A), ...)$. Also, given an orientation

of the transversal, each sequence has an inverse. Notice, the second table on page 10 did not include inverses, since they corresponded with the micronuclear sequences of the HPP orientations. As an observation, there are at least eight micronuclear sequence (unless they are complementary or assembly graph 11). For the other orientations Γ^- and Γ^{-R} there are eight new sequences, potentially.

5 Acknowledgments

References

[1] A. Angeleska, N.Jonoska, M. Saito, *DNA recombination through assembly graphs.* 2009, 3020-3027.

[2] A. Cavalcanti, T.H. Clarke, L. Landweber, *MDS_IES_DB: a database of macronuclear and micronuclear genes in spirotrichous ciliates. Nucleic Acids Research.* 33, (2005), 396398.

[3] J.Burns, E. Dolzhenko, N.Jonoska, T. Muche, M. Saito, *Four-Regular Graphs with Rigid Vertices Associated to DNA Recombination.* 2011, 1-13.

[4] M. Saito, *Tables of Assembly Graphs.* 2009, 1-4.