Enumeration of Irreducible Multiple-Occurrence Words

46th SEICCGTC

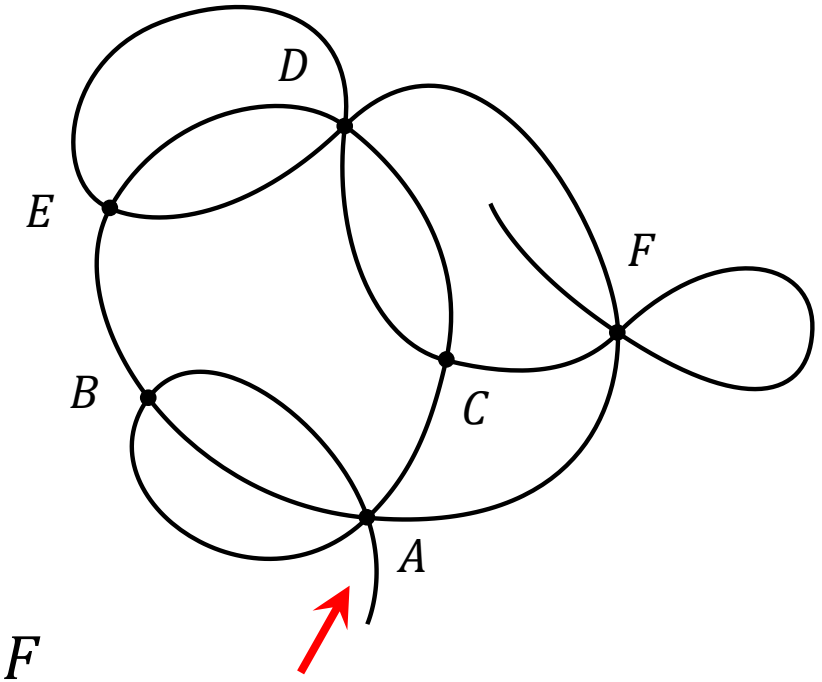
Jonathan Burns

University of South Florida

Princeton University

Smooth, Self-Intersecting Curves

Consider a smooth curve $L: [0,1] \to \mathbb{R}^3$ with a finite number of intersections:



Transverse Path:

$A\ B\ A\ C\ D\ E\ B\ A\ F\ D\ E\ D\ C\ F\ F$

Multi-Occurrence Words

Each letter occurs at least twice

- Example

Symbol:	A	B	A	A	C	B	C	C	A
Position:	1	2	3	4	5	6	7	8	9

Multi-Occurrence Words

Each letter occurs at least twice

- Example

Symbol:	**A** B **A** **A** C B C C **A**
Position:	**1** 2 **3** **4** 5 6 7 8 **9**

Partition: **A:** {1,3,4,9}, B: {2,6}, C: {5,7,8}

Definition

The r-associated Stirling numbers of the second kind, denoted $S_r(n,k)$, are the number of ways to form a partition of size n having k blocks with at least r elements.

Recursive relation:

$$S_r(n,k) = k\, S_r(n-1,k) + \binom{n-1}{r-1}\, S_r(n-r,k-1)$$

Fact

Let M_n be the set of multi-occurrence words of length n.

Then

$$|M_n| = \sum_{j=1}^{\lfloor n/2 \rfloor} S_2(n, j)\,.$$

Let K_n be the set of k-occurrence words of length kn. Then

$$|K_n| = \frac{(nk)!}{(n!)(k!)^n}\,.$$

Symmetric Words

A word w is **symmetric** if $w^R = w$ after a relabeling.

Example:

$$w = 1223143554$$

$$w^R = 4553413221$$

$$w_*^R = 1223143554$$

$1 \to 4,$
$2 \to 5,$
$3 \to 3,$
$4 \to 1,$
$5 \to 2$

Symmetric Partitions

A partition $\mathcal{P}$ of size n is **symmetric** if for each $i, j \in [n]$ belonging to a block of $\mathcal{P}$, the elements $(n - i + 1)$ and $(n - j + 1)$ also belong to a block of $\mathcal{P}$.

Example: $w = 1223143554$

$$\mathcal{P}_w = \{1,5\}, \{2,3\}, \{4,7\}, \{6,10\}, \{8,9\}$$
$$1 \quad 2 \quad 3 \quad 4 \quad 5$$

Symmetric Multi-Occurrence Words

Proposition: Let P_n be the set of symmetric multi-occurrence words of length n. Then for $k \geq 1$,

$$\left|P_{2(k+1)}\right| = |P_{2k}| + \sum_{j=1}^{k} \binom{k}{j} (2^j + 1)\left|P_{2(k-j)}\right|$$

$$\left|P_{2(k+1)+1}\right| = \sum_{j=1}^{k} \binom{k}{j} \left|P_{2(k-j)}\right|$$

with the initial conditions $|P_0| = |P_1| = |P_2| = |P_3| = 1$.

Symmetric k-Occurrence Words

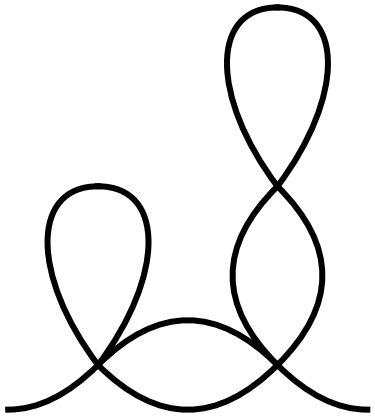
Proposition: Let $L_{n,k}$ be the set of symmetric k-occurrence words of length kn. Then for $k = 2j$,

$$L_{n,2j} = \binom{nj-1}{j-1} L_{n-1,2j} + 2^{2j-1} \binom{nj-1}{j-1} L_{n-2,2j}$$

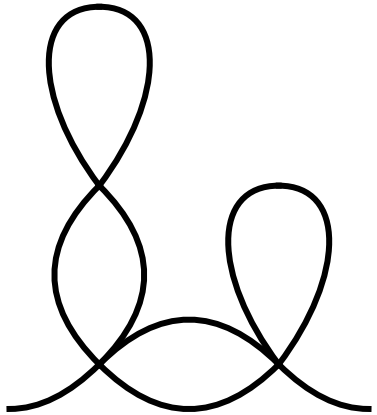
and for $k = 2j+1$,

$$L_{n,2j+1} = \begin{cases} \left(\binom{(nj-1)+\binom{n}{2}}{2j} \right) L_{n-1,2j+1} + 2^{2j} \left(\binom{(nj-1)+\binom{n}{2}}{2j} \right) L_{n-2,2j+1}, & \text{if } n \text{ is odd} \\ 2^{2j} \left(\binom{(nj-1)+\binom{n}{2}}{2j} \right) L_{n-2,2j+1}, & \text{if } n \text{ is even} \end{cases}$$

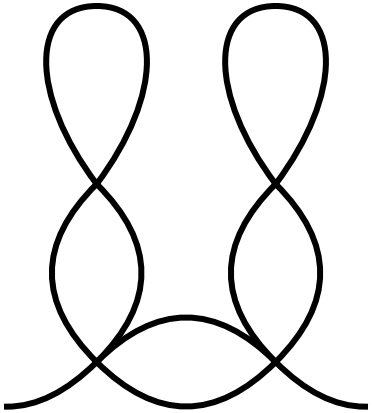
Multi-Occurrence Words and Non-Isomorphic Self-Intersection Curves



112332



122133



12213443

$$\#\text{Non-Iso. Curves} = \frac{1}{2}(\#\text{ Multi} + \#\text{Symmetric Multi})$$

Irreducible Multi-Occurrence Words

Reducible:

$w = u\,v \qquad u, v \in M_n$ **Ex:** 121<u>34343</u>

Weakly Irreducible (I_n):

$w = u_1 v\,u_2 \quad u_1 u_2,\ v \in M_n$ **Ex:** 123<u>4545</u>213

Strongly Irreducible (S_n):

w is not weakly irreducible **Ex:** 1213233213

Multi-Occurrence Words

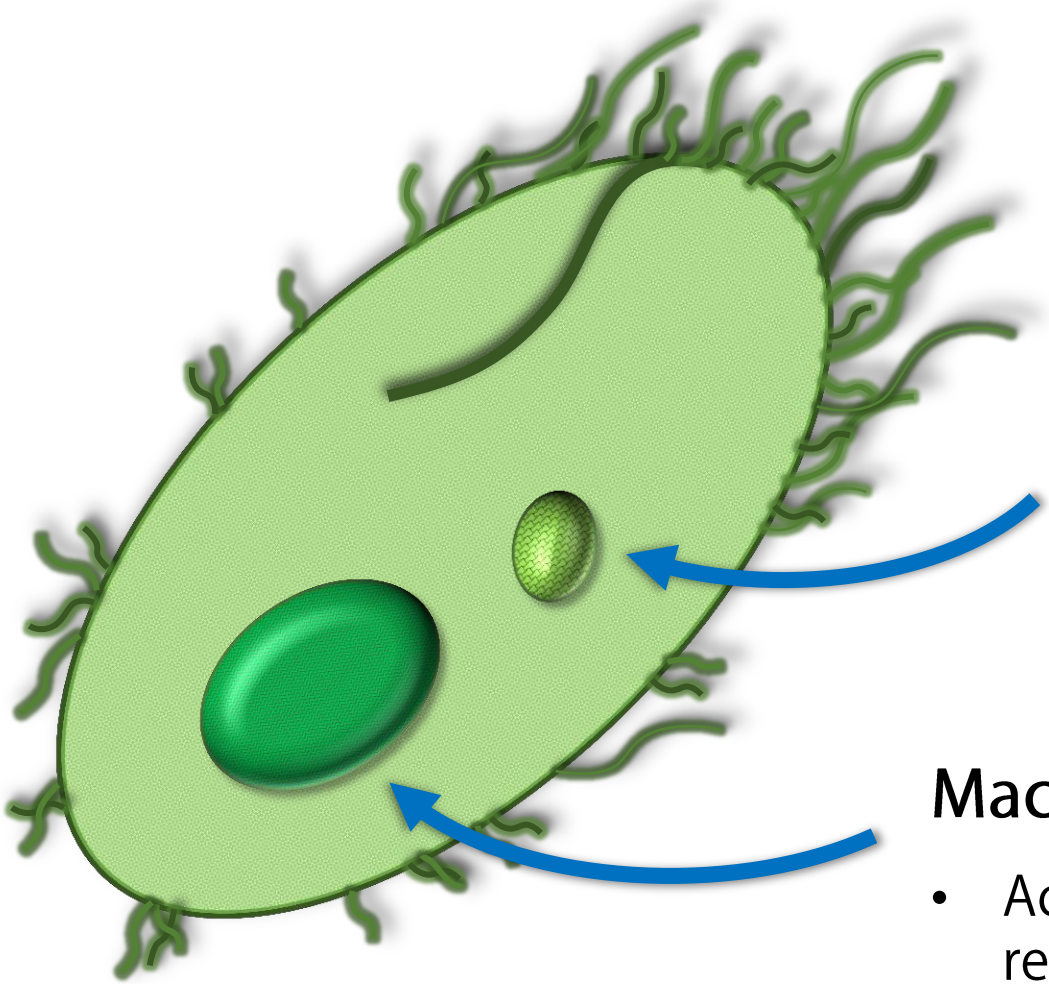
n	All	Symmetric	Weakly Irreducible	Strongly Irreducible
2	1	1	1	1
3	1	1	1	1
4	4	4	3	2
5	11	3	9	6
6	41	15	33	21
7	162	16	135	85
8	715	75	609	385
9	3,425	89	2,985	1,907
10	17,722	428	15,747	10,205
OEIS	A000296	A086365[1]	A098742	A099947

[1] even n only

Non-Isomorphic, Linear Self-Intersection Curves

n	All	Weakly Irreducible	Strongly Irreducible
2	1	1	1
3	1	1	1
4	4	3	2
5	7	6	4
6	28	23	14
7	89	75	47
8	395	338	210
9	1,757	1,535	979
10	9,075	8,072	5,208

Biological Motivation



Ciliated protozoans

- Single-celled
- Two types of nucleus

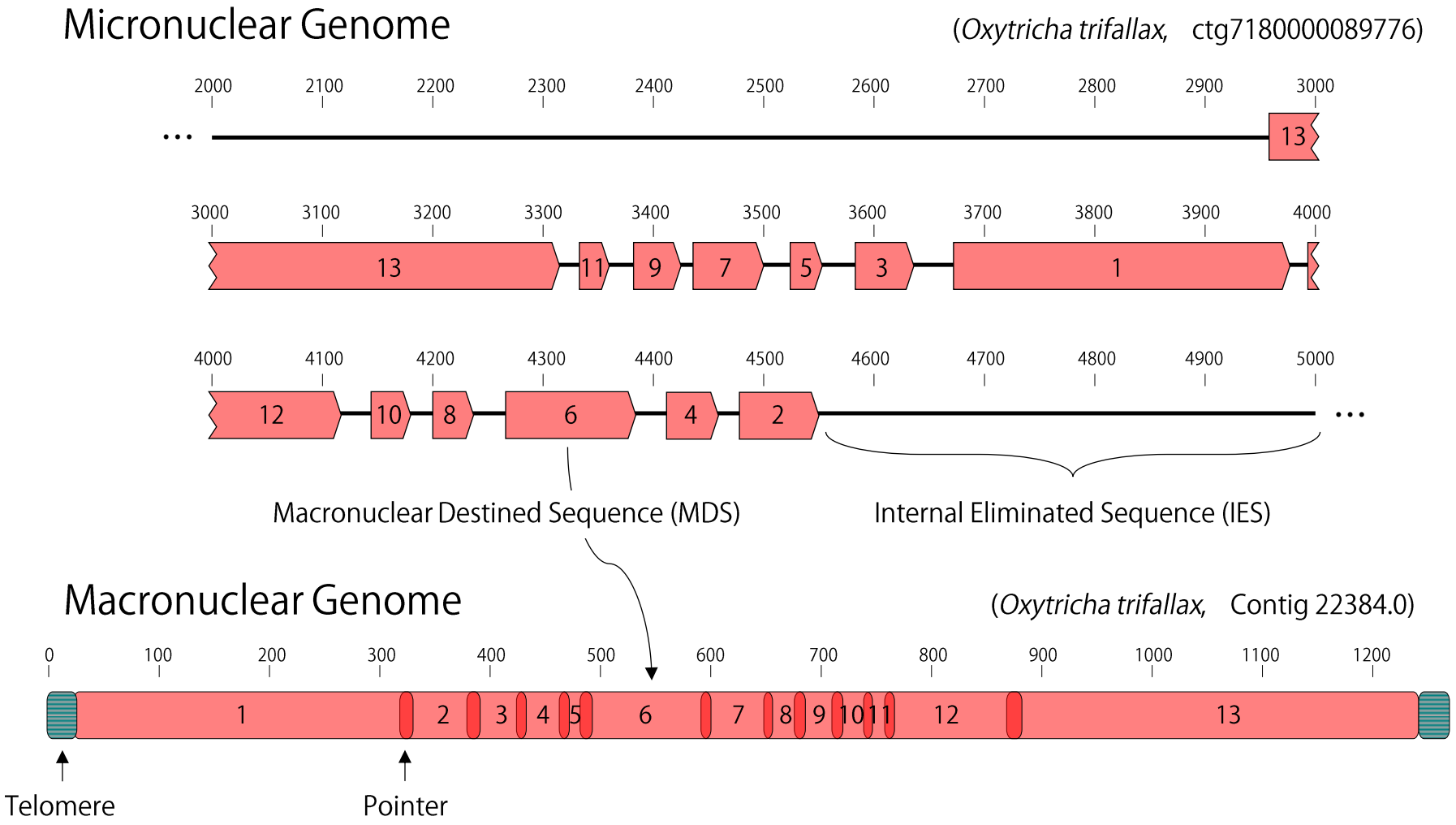
Micronucleus (MIC)

- Backup copy of the macronucleus

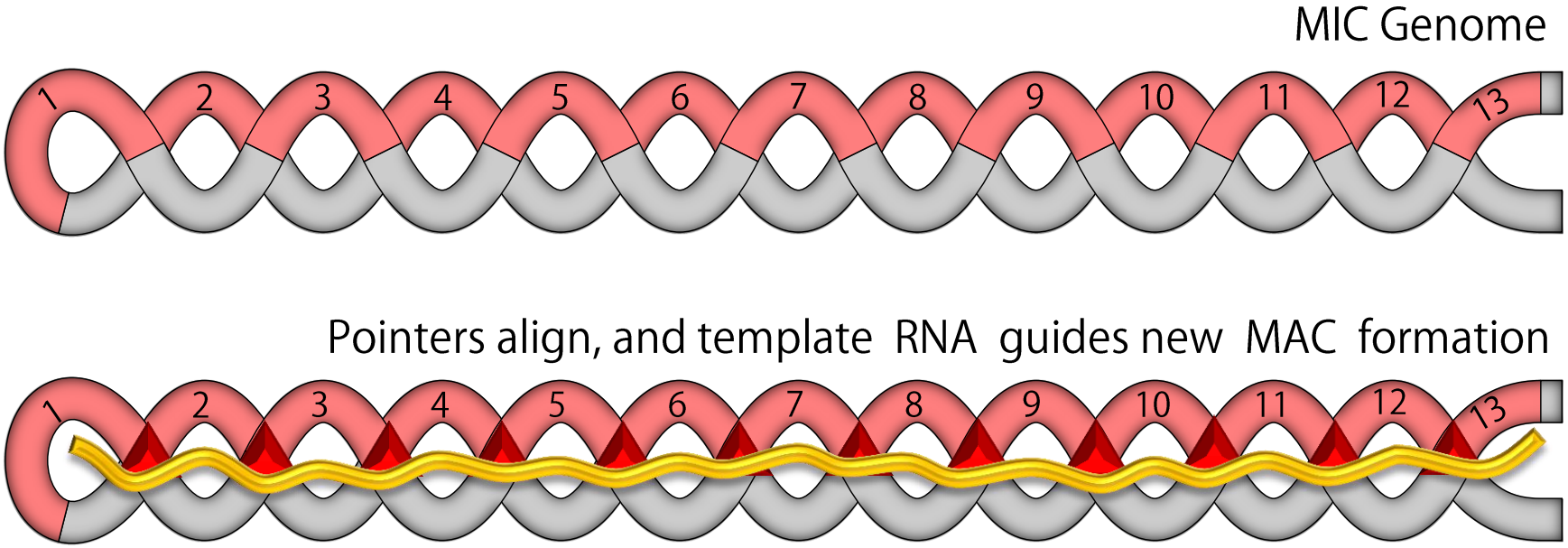
Macronucleus (MAC)

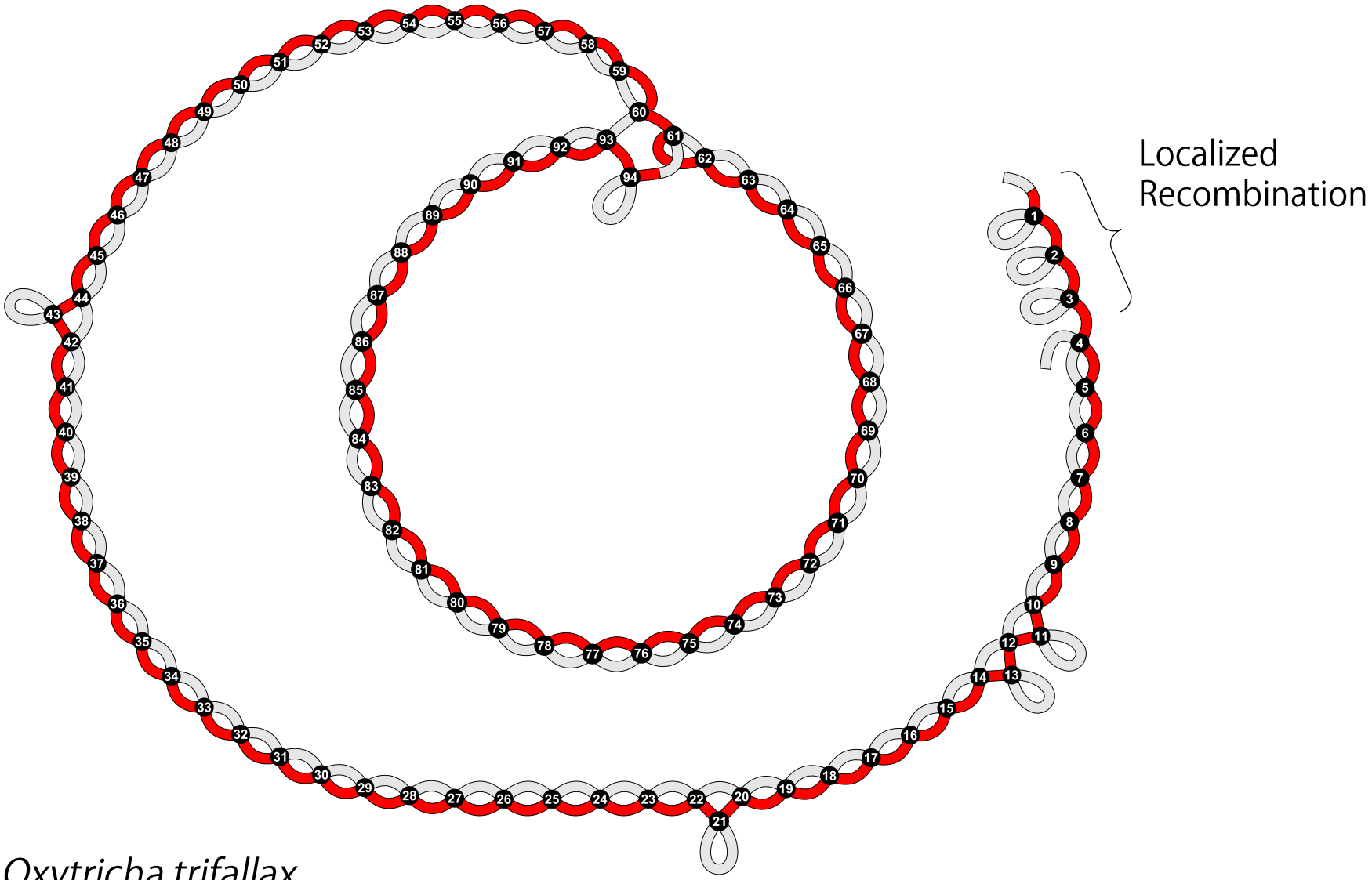
- Actively transcribes RNA for regular cell functions

MAC and MIC have different genomes



Assembly Graph Model (Angeleska, et al 2007)

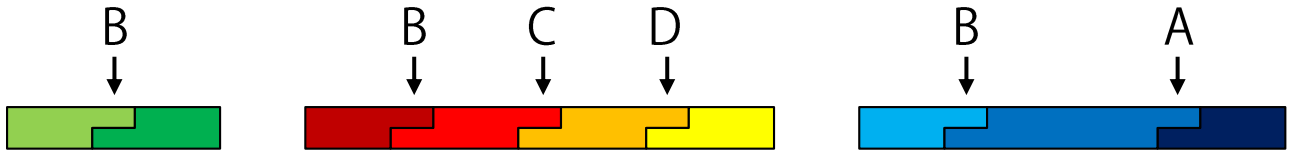




Oxytricha trifallax
(ctg7180000067914 / Contig45.1)

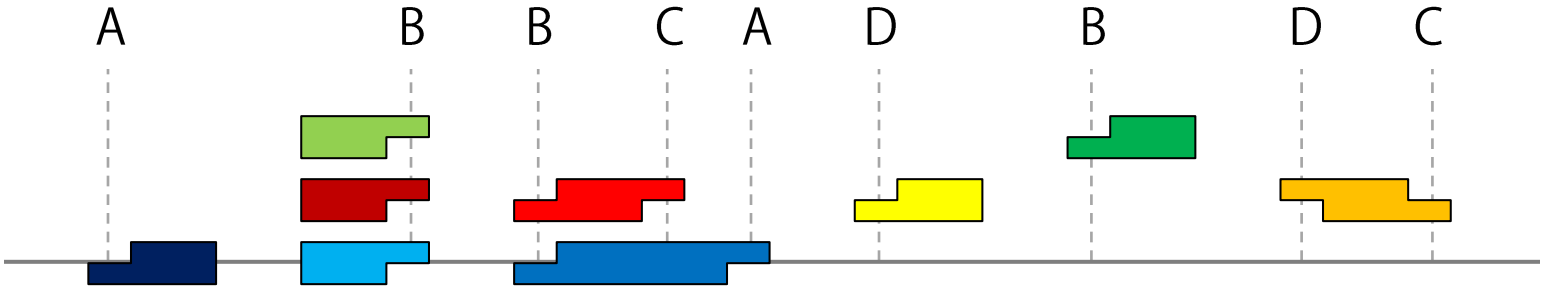
Generalized Assembly Graph Model

- Identify pointers by the overlaps in MAC sequence



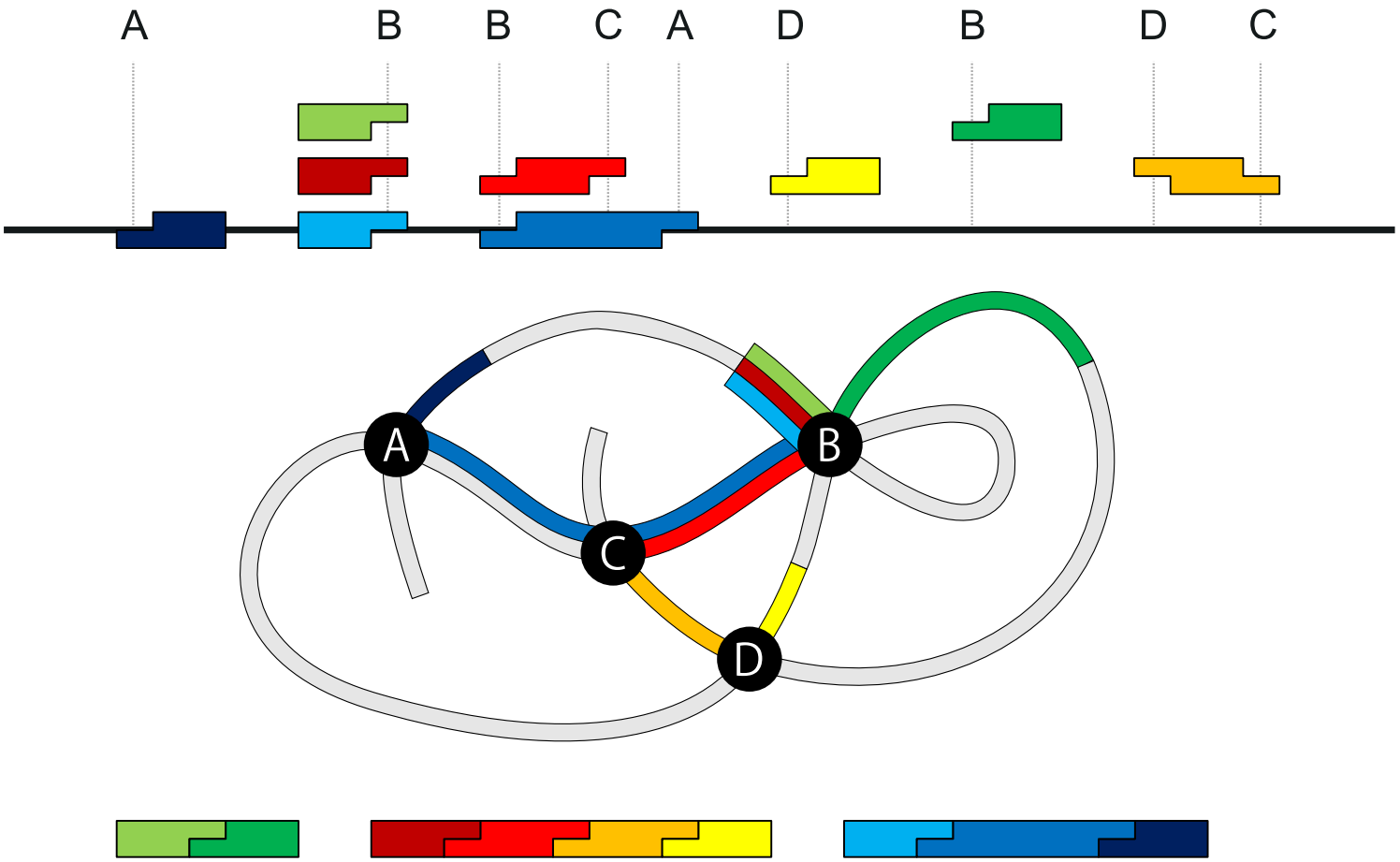
(Possible MAC chromosomes)

- Assign two pointers the same symbol if they overlap in either MAC or MIC



(MIC Genome)

Both MIC and MAC as Assembly Graph



Math-Bio Lab
University of South Florida

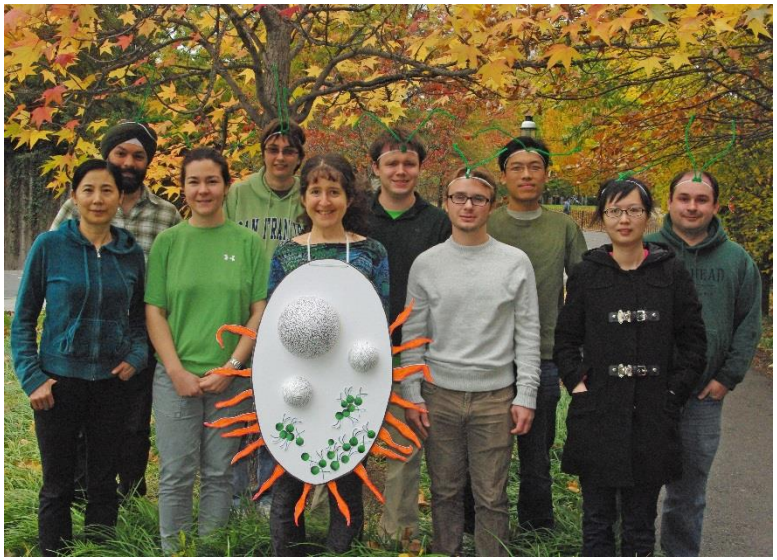
Lab Heads
- Natasha Jonoska
- Masahico Saito

Members
- Daniel Cruz
- Denys Kukushkin
- Daria Karpenko
- Rick Wallace

Landweber Lab
Princeton University

Lab Head
- Laura Landweber

Members
- Leslie Beh
- Derek Clay
- Rick Miller
- Talya Yerlici
- Xiao Chen
- Jaspreet Khurana
- Jingmei Wang



Funded by: National Institutes of Health – Grant #R01 GM109459-01

Thanks!