

**Computability in Europe 2014**

(session on Bio-inspired computation)

Budapest, Hungary

June 2014

**Graph polynomials motivated by  
Gene Rearrangements in Ciliates**

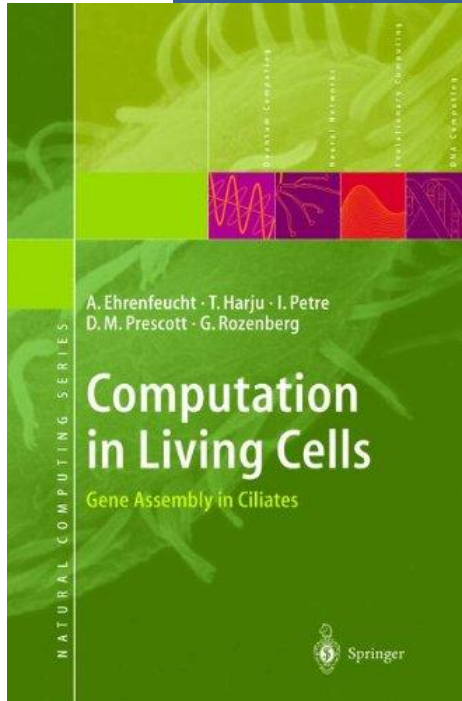
Hendrik Jan Hoogeboom

Leiden NL

joint with

Robert Brijder

Hasselt B



## Computation in Living Cells

– Gene Assembly in Ciliates

A.Ehrenfeucht, T.Harju, I.Petre,

D.M.Prescott, G.Rozenberg

Springer, 2004

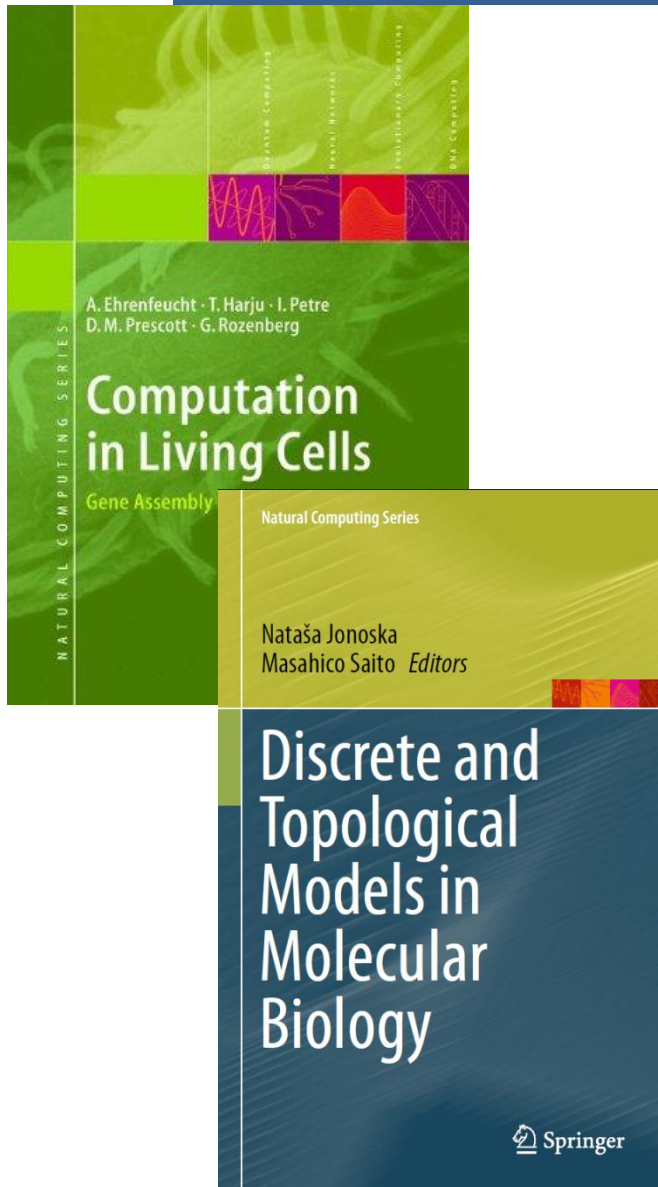
“This indicates that, in principle, these unicellular organisms may have the capacity to perform at least any computation carried out by an electronic computer”

[L. Landweber, L. Kari: Universal Molecular Computation in Ciliates, 2002]

“The way that ciliates transform genes from their micronuclear (storage) form into their macronuclear (expression) form constitutes a very interesting case of ‘DNA computing *in vivo*’ ”

[Ehrenfeucht, Prescott, Rozenberg: Computational Aspects of Gene (Un)Scrambling in Ciliates, 2002]





## Computation in Living Cells

– Gene Assembly in Ciliates

A.Ehrenfeucht, T.Harju, I.Petre,

D.M.Prescott, G.Rozenberg

Springer, 2004

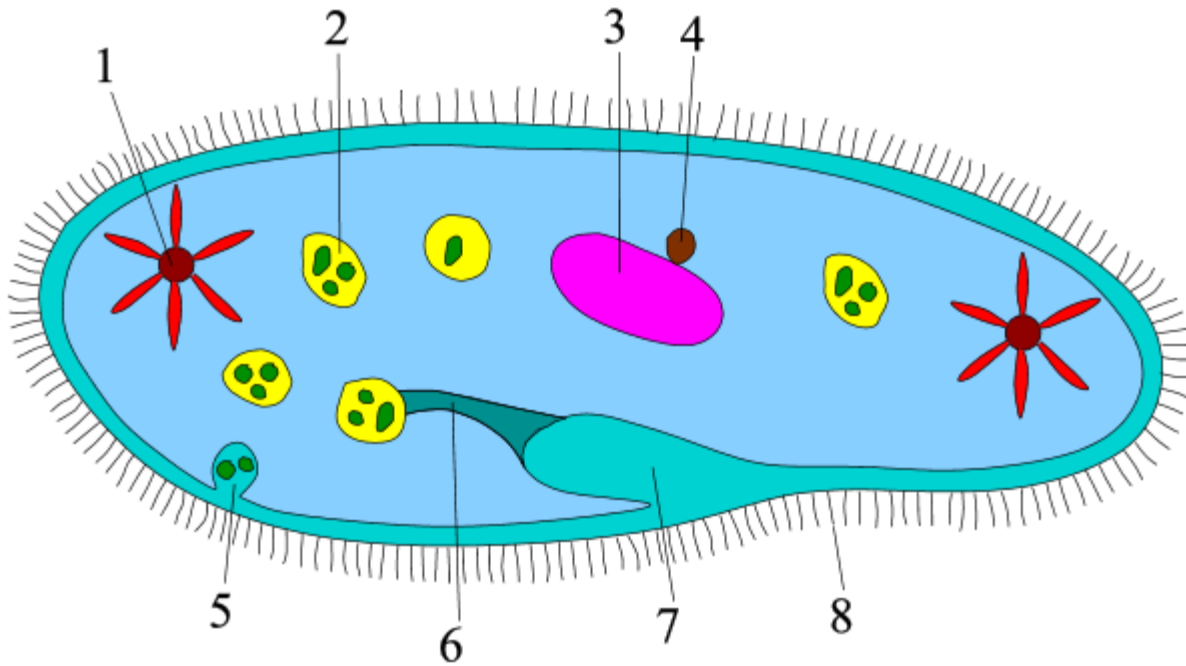
R. Brijder, H.J. Hoogeboom

The Algebra of Gene Assembly in Ciliates

In: *Discrete and Topological Models in  
Molecular Biology* (N. Jonoska, M. Saito, eds.),

Springer, 289–307, 2014

# ciliates: micro and macro



cell structure:

- 3. macronucleus
- 4. micronucleus
- 8. cilium

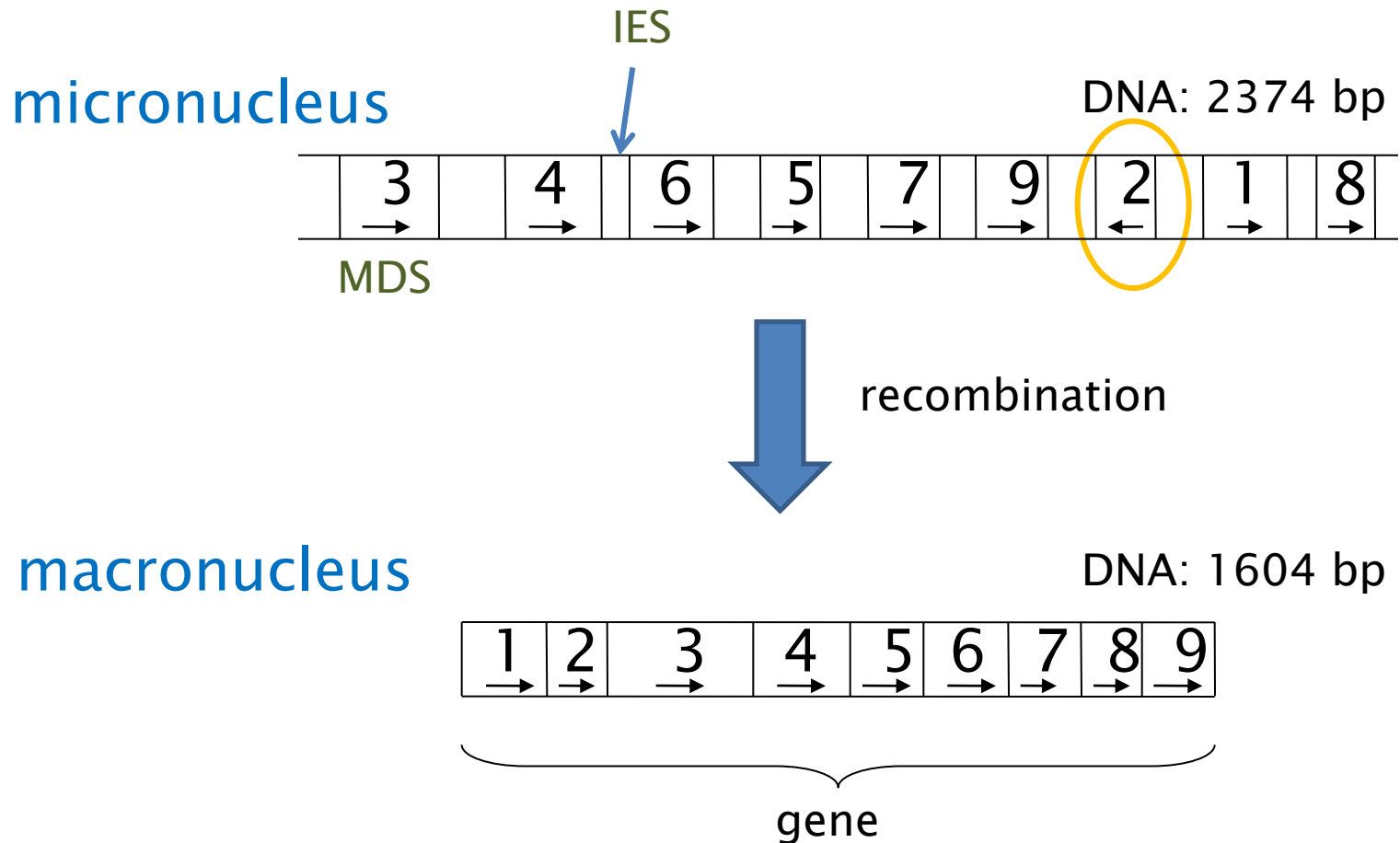
Two different sorts of nuclei:

- **micronucleus** (small, diploid, reproduction)
- **macronucleus** (large, polyploid, cell regulation)

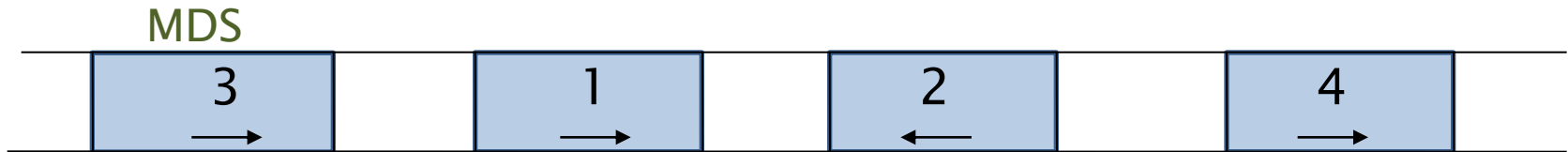
gene assembly: **micro**nucleus → **macro**nucleus

# from micro to macro

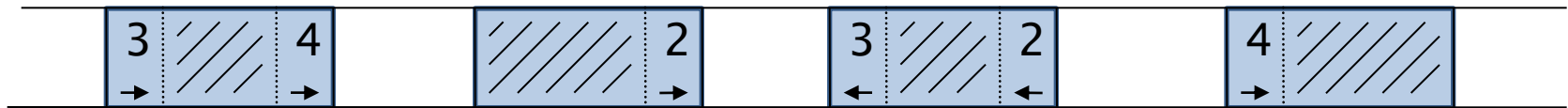
[http://oxytricha.princeton.edu/cgi-bin/get\\_MDS\\_IES\\_Info.cgi?num=38](http://oxytricha.princeton.edu/cgi-bin/get_MDS_IES_Info.cgi?num=38)



Greslin, Prescott et al. Reordering of nine exons is necessary to form a functional actin gene in *Oxytricha nova*. PNAS 86, 6264–6268, Aug 1989.

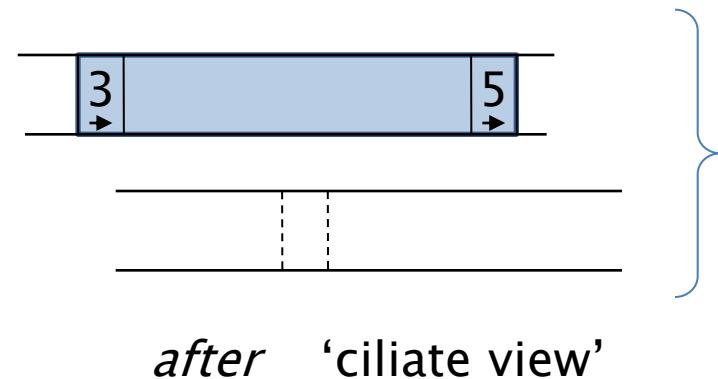
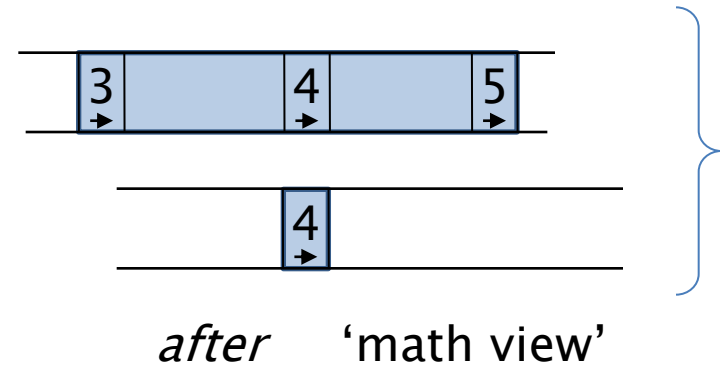
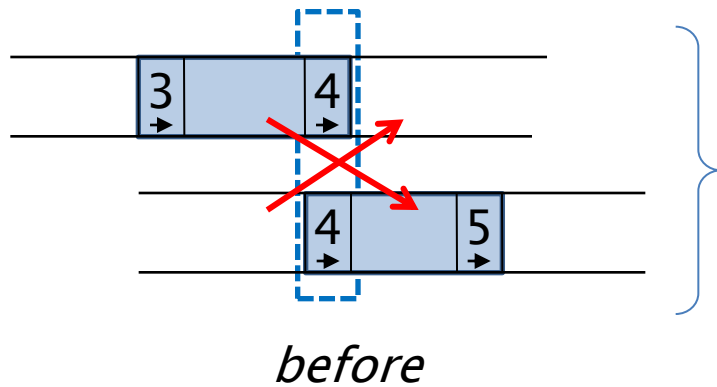


pointers – overlapping segments



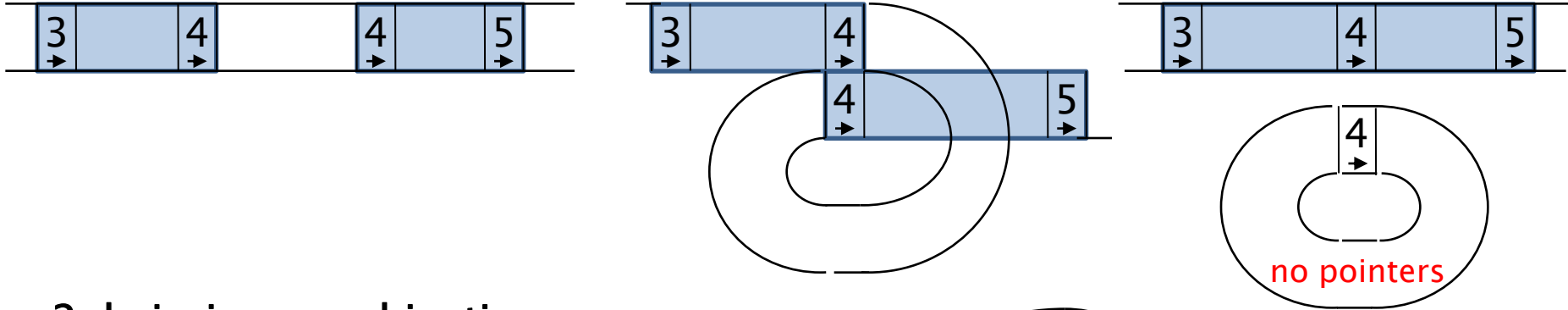
e.g., pointer 5 of actin gene: 13 bp

$rc_4$  recombination on pointer 4 'generic'

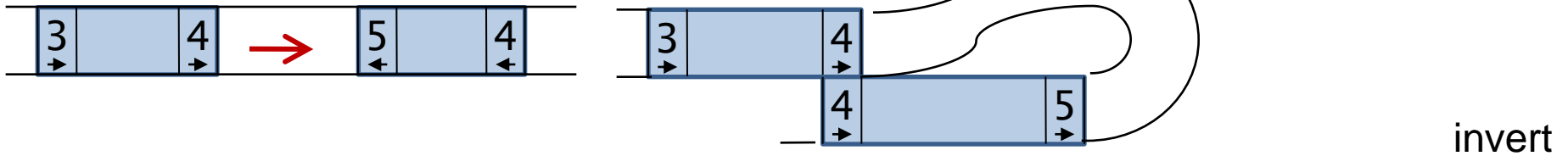


# recombination on pointers

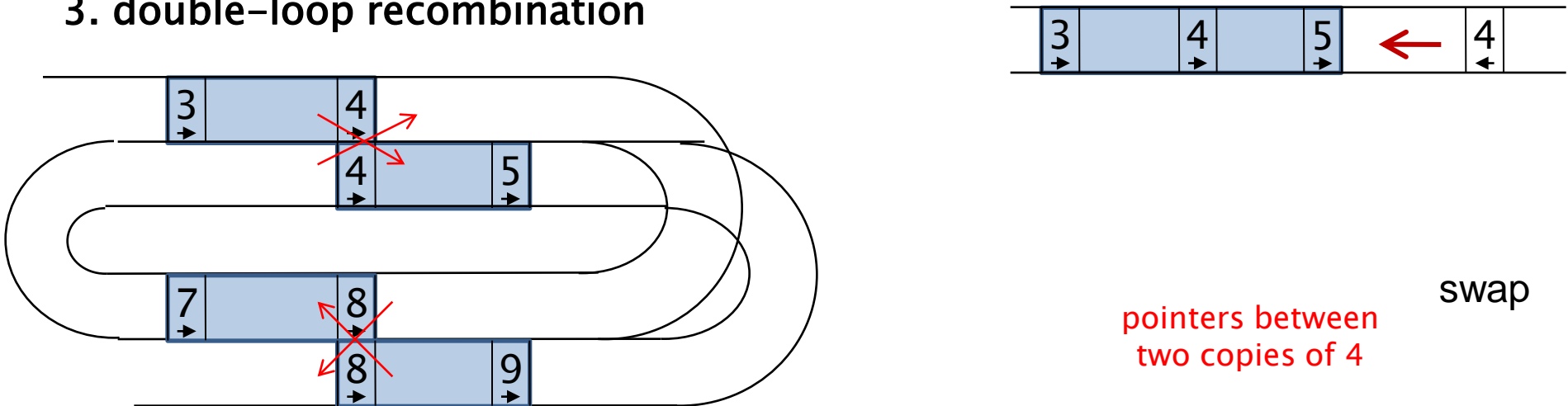
## 1. loop recombination



## 2. hairpin recombination

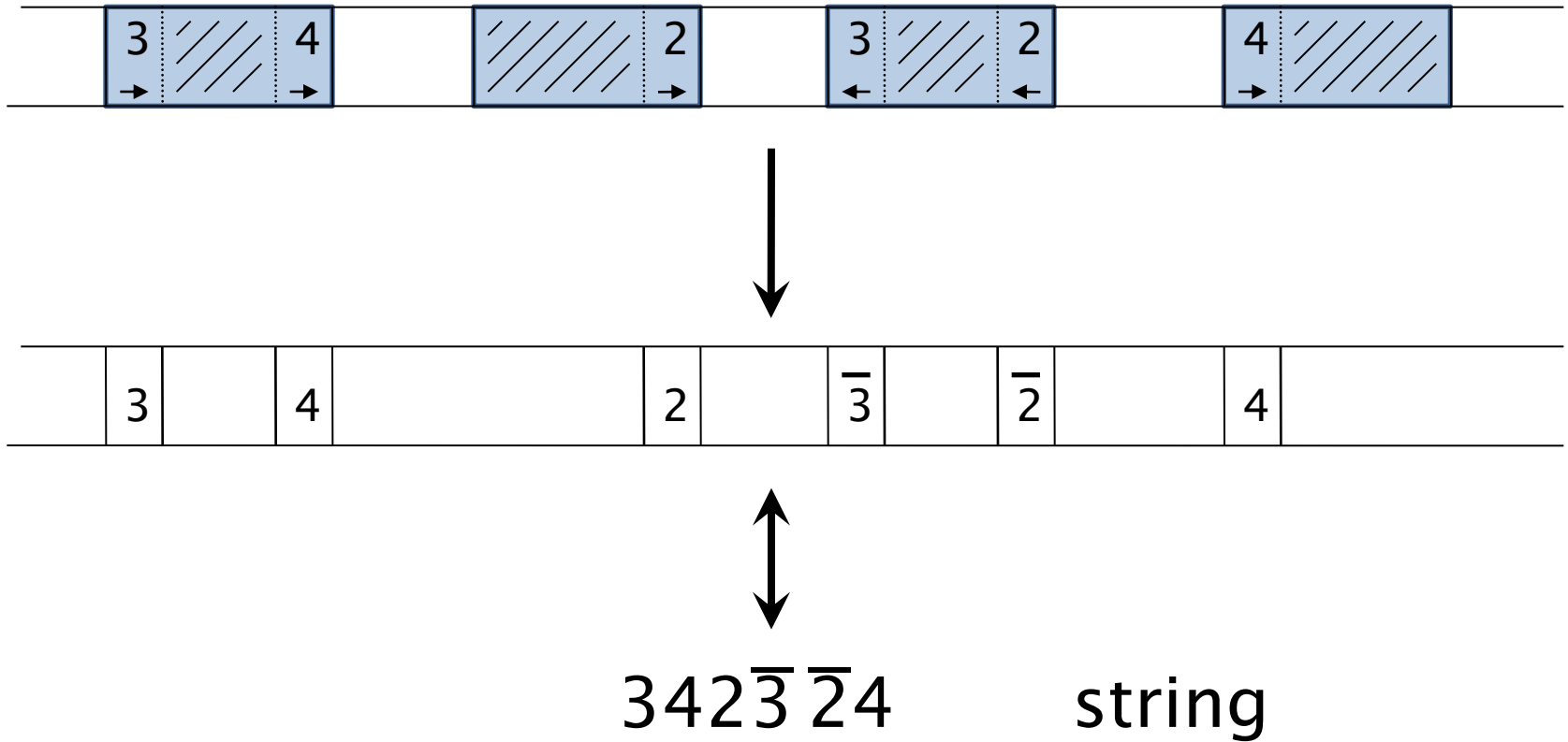


## 3. double-loop recombination





# abstraction: pointers



# string pointer reduction systems

no rearrangement  
excision circular molecule

$$\text{spr}_p( u_1 p p u_2 ) = u_1 p p u_2$$

split



$$\text{spr}_p( u_1 p u_2 \bar{p} u_3 ) = u_1 p \bar{u}_2 \bar{p} u_3$$

invert

$$\text{spr}_{p,q}( u_1 p u_2 q u_3 p u_4 q u_5 ) = u_1 p u_4 q u_3 p u_2 q u_5$$

swap

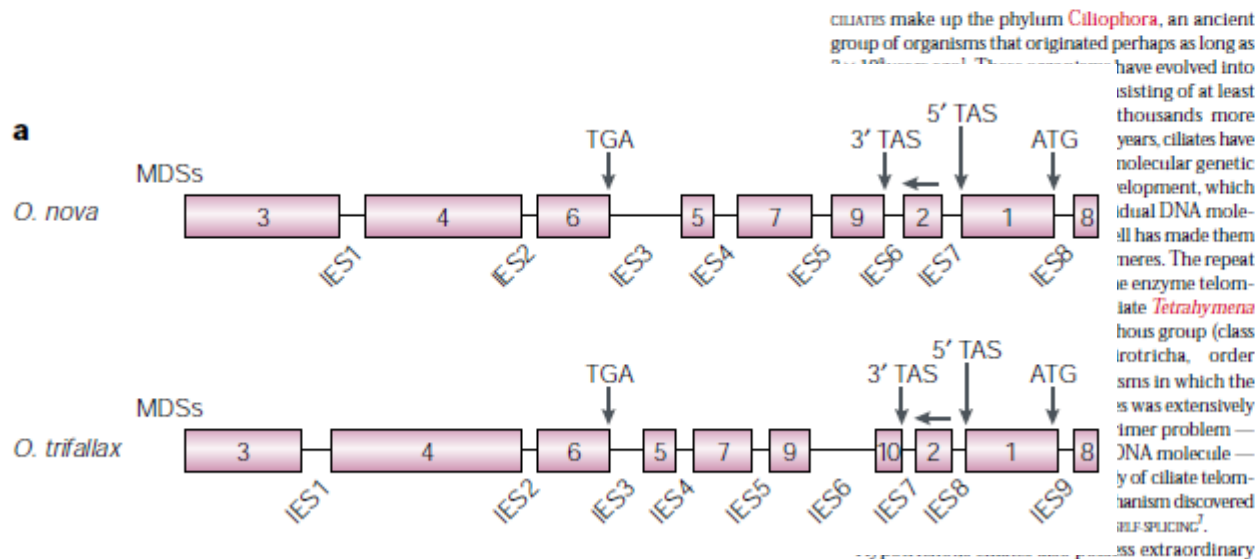
in “ciliate view” the **p** pointers disappear

# look at the “gymnastics”

## GENOME GYMNASTICS: UNIQUE MODES OF DNA EVOLUTION AND PROCESSING IN CILIATES

David M. Prescott

In some ciliates, the DNA sequences of the germline genomes have been profoundly modified during evolution, providing unprecedented examples of germline DNA malleability. Although the significance of the modifications and malleability is unclear, they may reflect the evolution of mechanisms that facilitate evolution. Because of the modifications, these ciliates must perform remarkable feats of cutting, splicing, rearrangement and elimination of DNA sequences to convert the chromosomal DNA in the germline genome (micronuclear genome) into gene-sized DNA molecules in the somatic genome (macronuclear genome). How these manipulations of DNA are guided and carried out is largely unknown. However, the organization and manipulation of ciliate DNA sequences are new phenomena that expand a general appreciation for the flexibility of DNA in evolution and development.



CILIATES make up the phylum *Ciliophora*, an ancient group of organisms that originated perhaps as long as 2 billion years ago. Over the course of their evolution, ciliates have evolved into a diverse group consisting of at least thousands more years, ciliates have molecular genetic development, which ideal DNA molecule has made them meres. The repeat enzyme telomerase (class I) is extensively used in these organisms in which the telomeric DNA was extensively modified. This is a major problem — DNA molecule — of ciliate telomerase mechanism discovered in 1982. This extraordinary

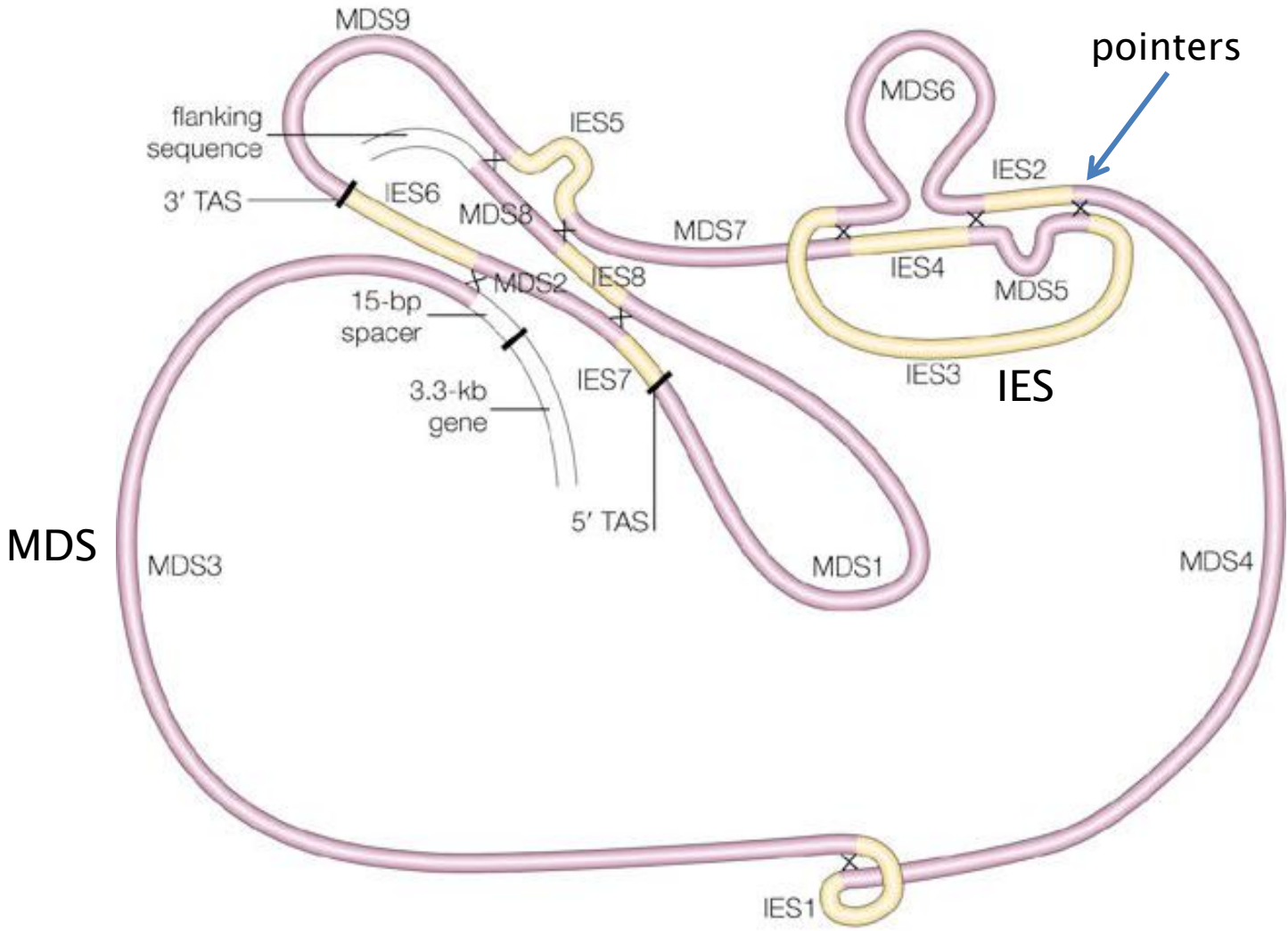
organizational features in their micronuclear and macronuclear genomes and carry out extensive developmental manipulations of their genomic DNA. These DNA phenomena, which considerably expand our view of genetic versatility, are the subject of this review.

### Germline and somatic nuclear DNA

A single ciliate organism has two kinds of nuclei: a germline diploid nucleus (micronucleus) and a somatic nucleus (macronucleus). The number of nuclei per cell varies in different ciliates — *Oxytricha trifallax* has two micronuclei and two macronuclei, whereas *Oxytricha nova* has four micronuclei and two macronuclei (FIG. 1). The micronucleus undergoes meiosis during cell mating, haploid micronuclei are exchanged between the two cells in a mating pair, and a new diploid micronucleus is formed in each cell (FIG. 1). The mating cells separate, and a macronucleus develops from a copy of the new diploid micronucleus. Simultaneously, unused haploid micronuclei and the old macronucleus are destroyed.

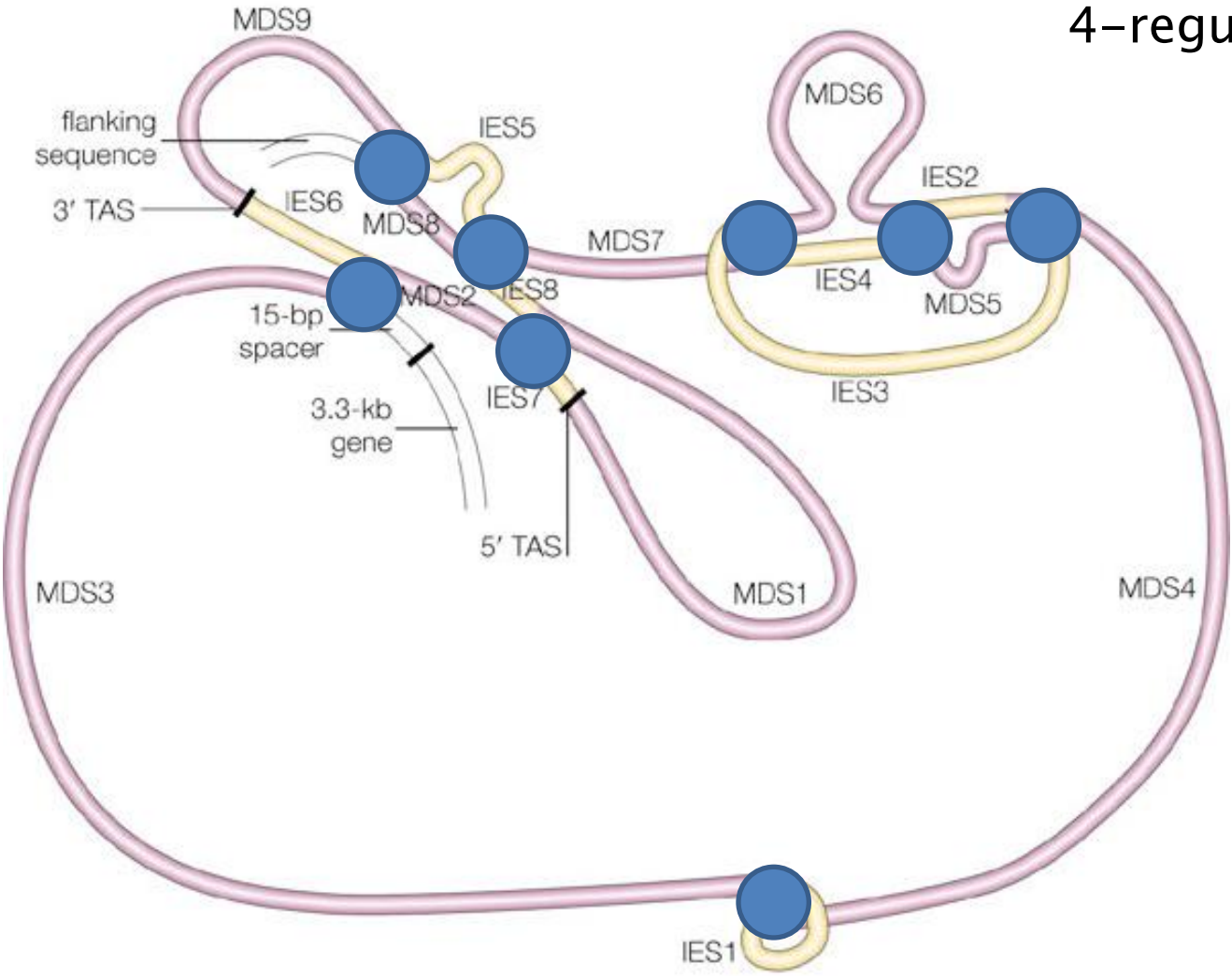
Micronuclear DNA consists of extremely long molecules typical of eukaryotic chromosomes. Genes are scattered along this DNA and are separated by large

# genome gymnastics

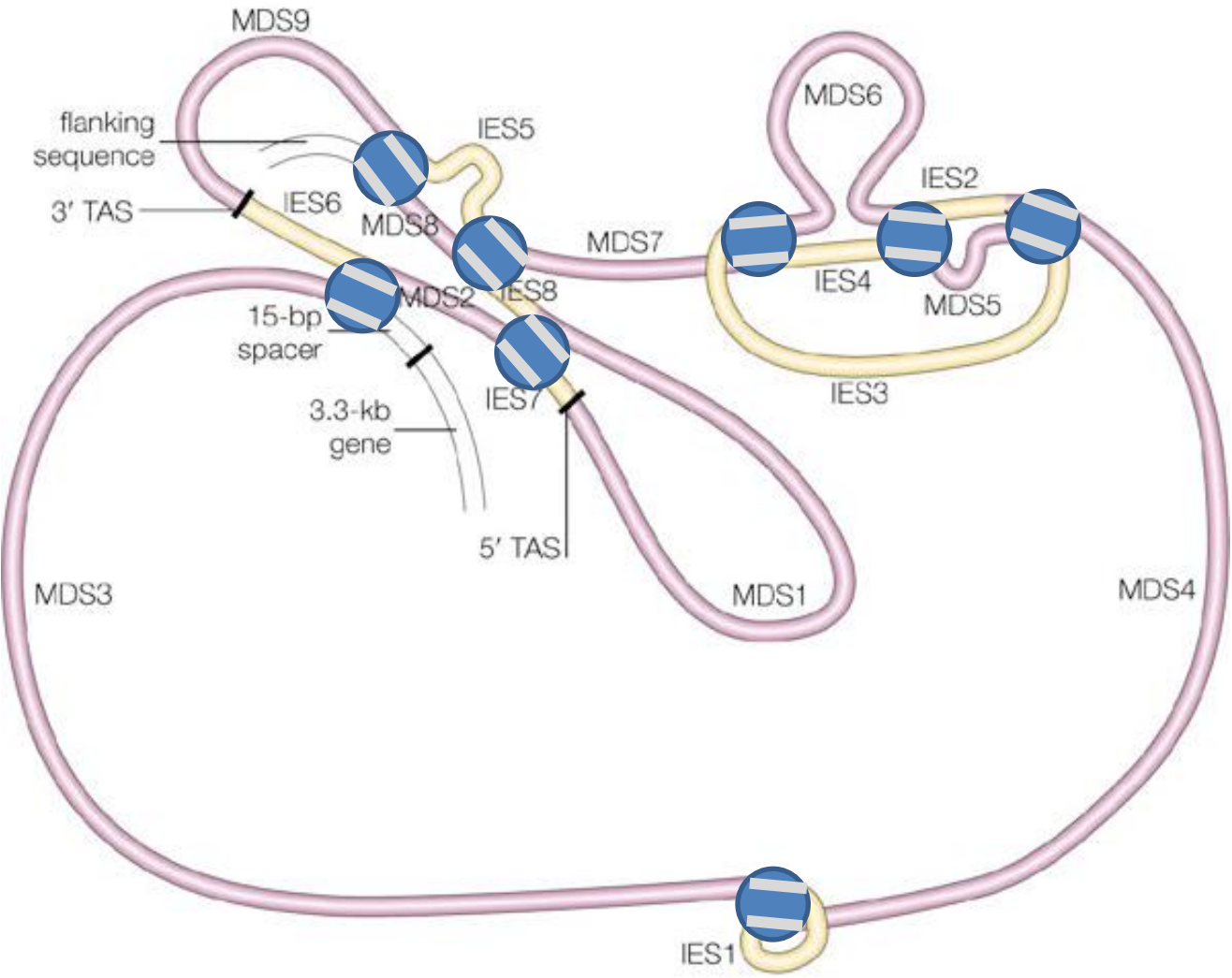


# genome gymnastics

4-regular graph

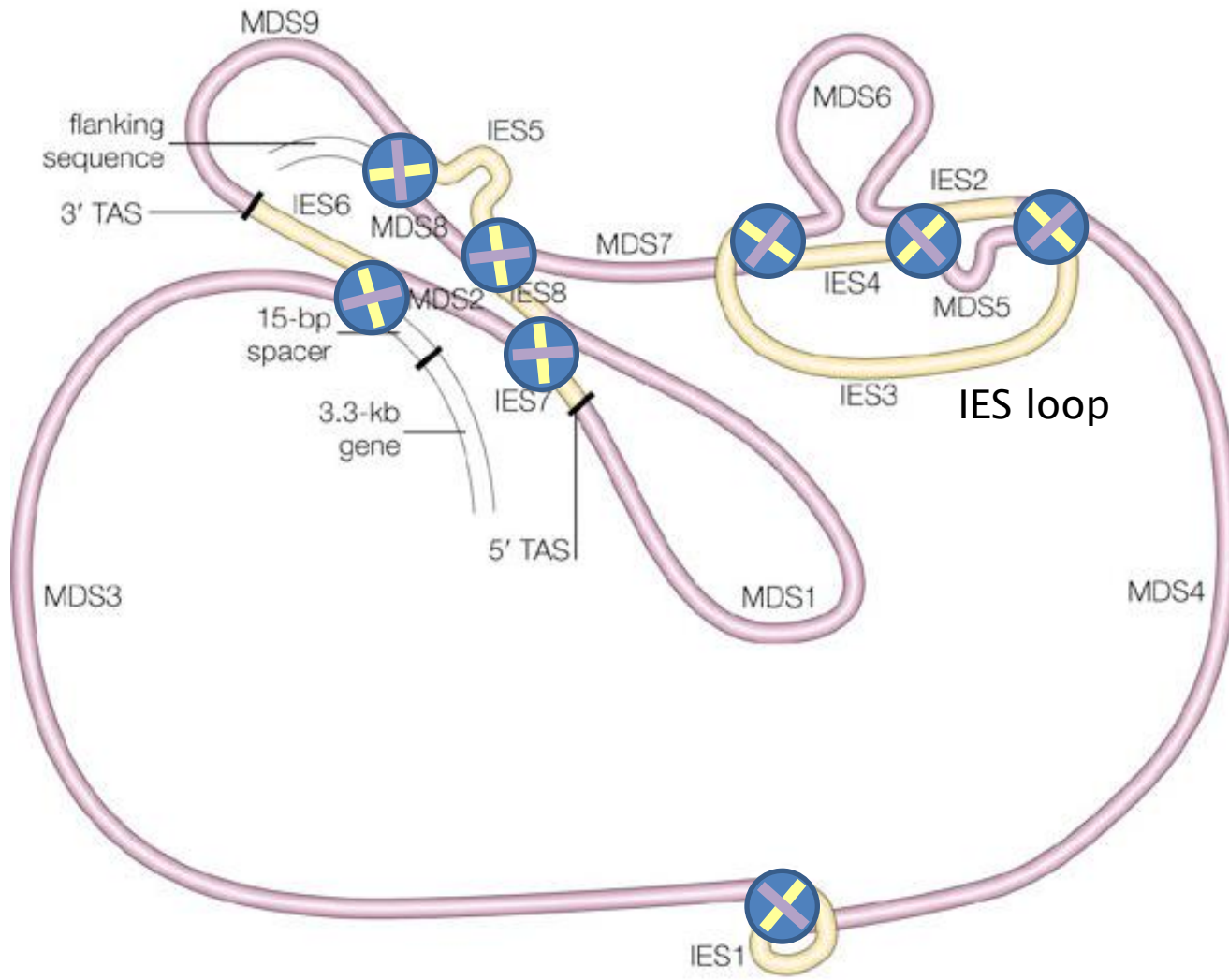


# genome gymnastics



**MIC**  
MDS and IES  
alternate





MAC

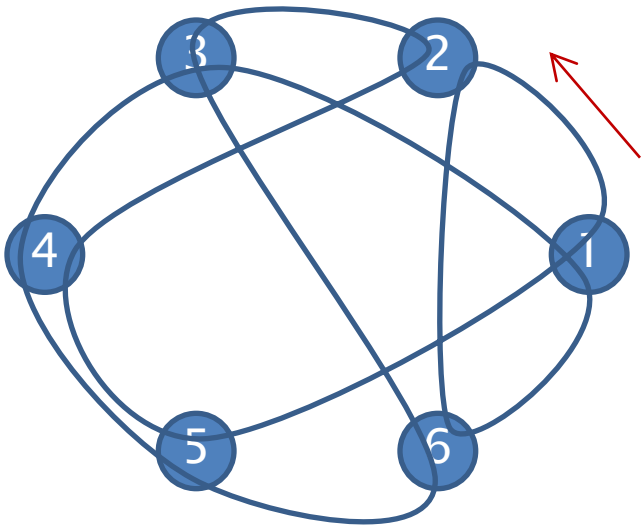
MDS and IES separated

How to study intramolecular gene assembly?

# Eulerian circuits in 4-regular graphs

string (without bars) = Euler circuit

126134563245



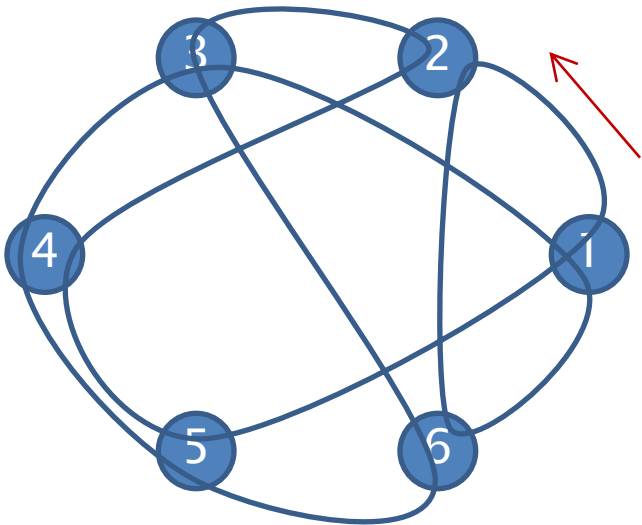
4-regular graph + Euler circuit

# Eulerian circuits in 4-regular graphs

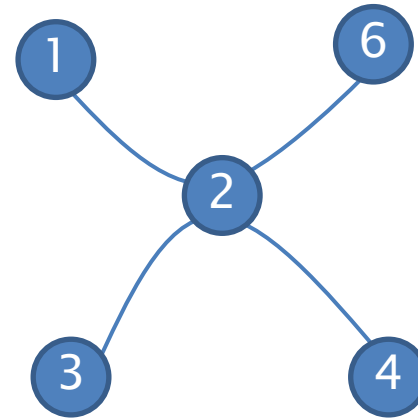
string (without bars) = Euler circuit

126134563245

1-2-6 & 3-2-4



three transitions at 2



1-2-4 & 3-2-6

1-2-3 & 6-2-4

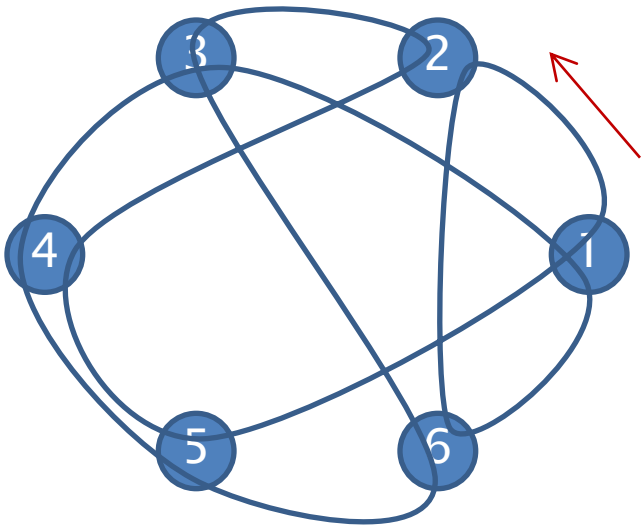
4-regular graph + Euler circuit

# Eulerian circuits in 4-regular graphs

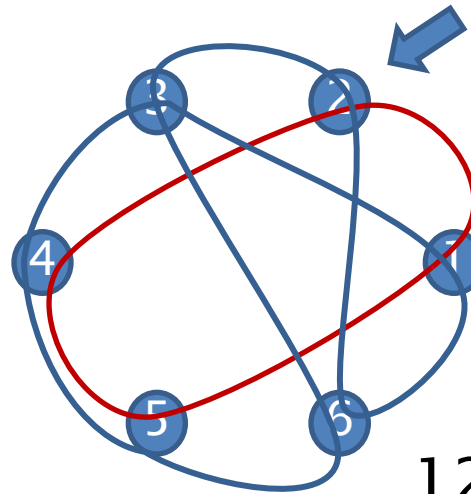
string (without bars) = Euler circuit

126134563245

1-2-6 & 3-2-4



4-regular graph + Euler circuit

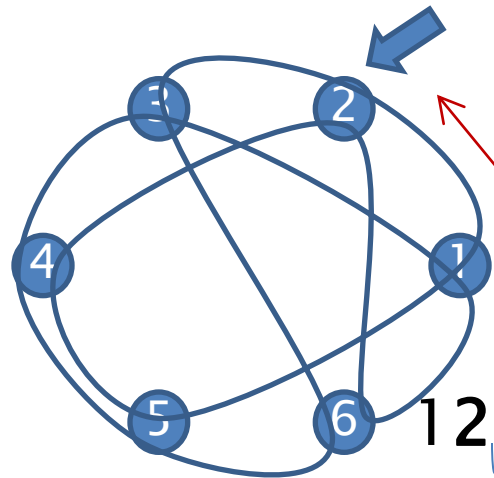


1-2-4 & 3-2-6

split

1245, 61345632

bars indicate whether **split** or **invert**  
(MDS to MDS)



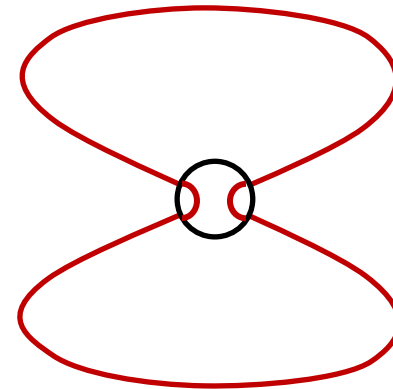
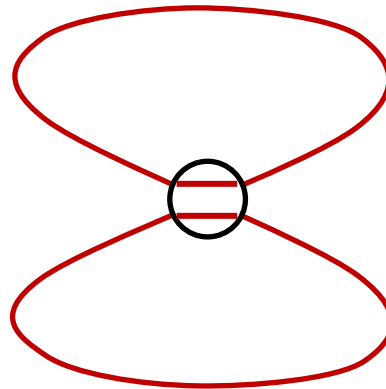
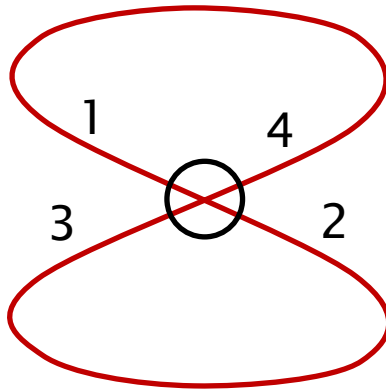
1-2-3 & 6-2-4

123654316245

invert

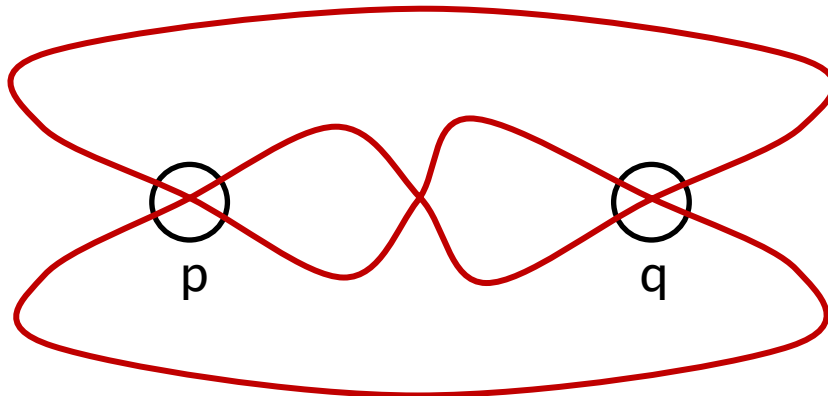
# reordering at vertices

Kotzig 1966



split

invert



*interlaced* splits: swap

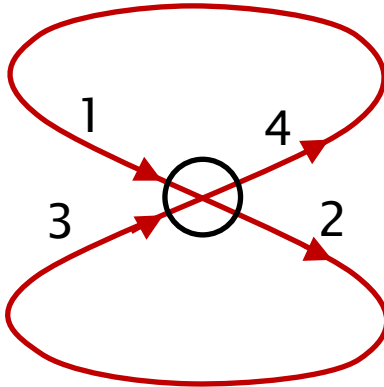
p ... q ... p ... q ...  
↑                    ↑

# reordering at vertices

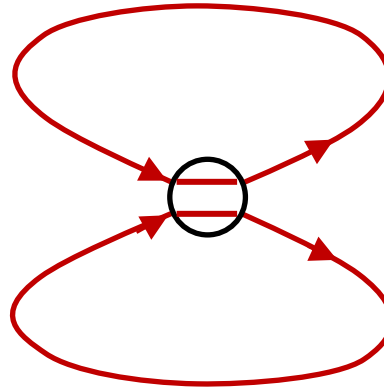
transition ...

Kotzig 1966

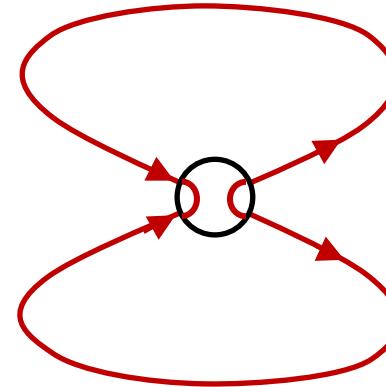
follows cycle



consistent

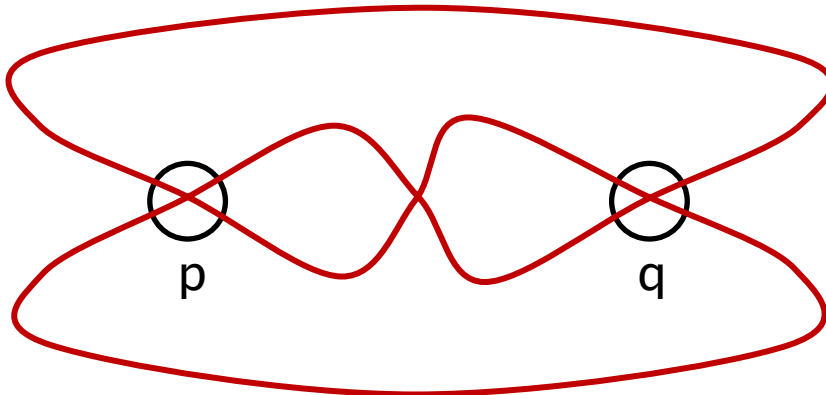


inconsistent



split

invert



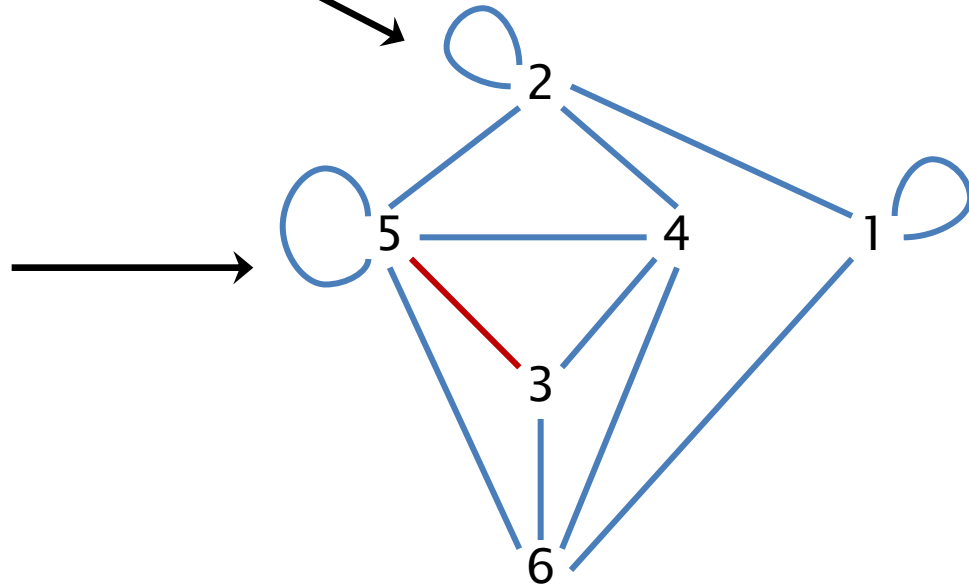
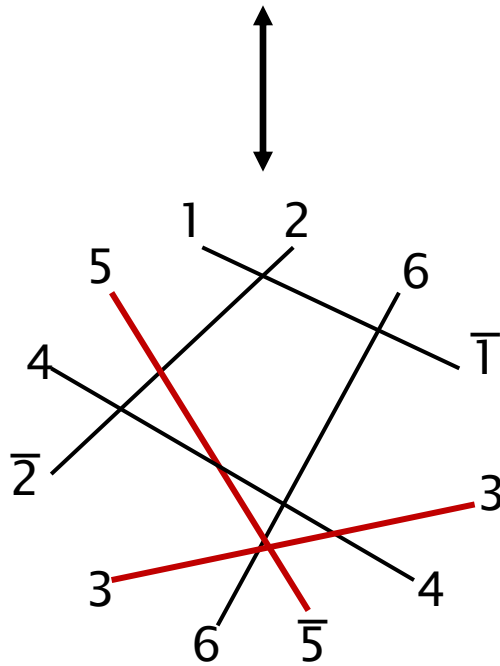
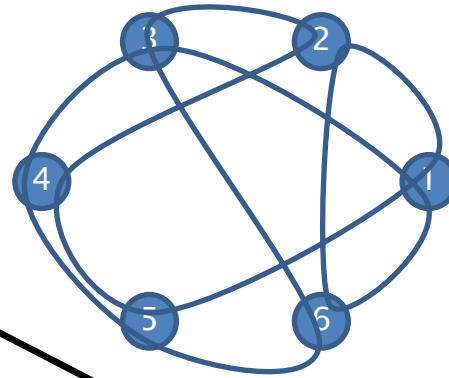
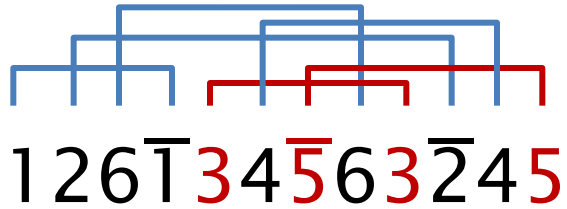
*interlaced* splits: swap

p ... q ... p ... q ...  
↑            ↑



# interlace / overlap graph

circle / interval graph

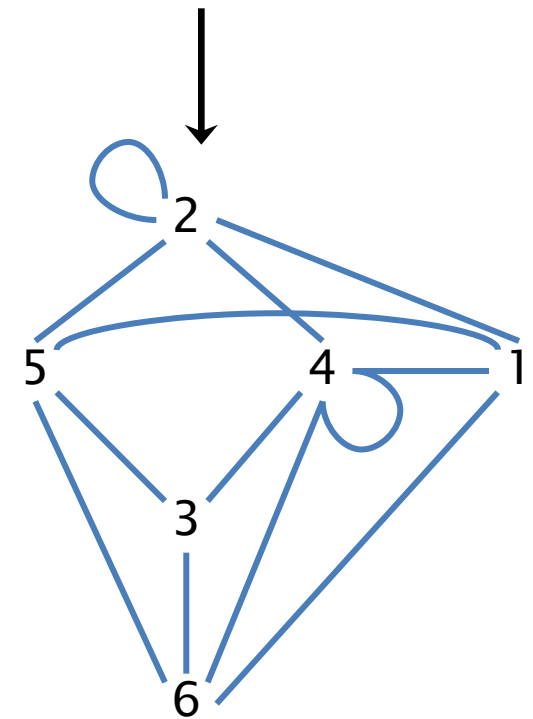
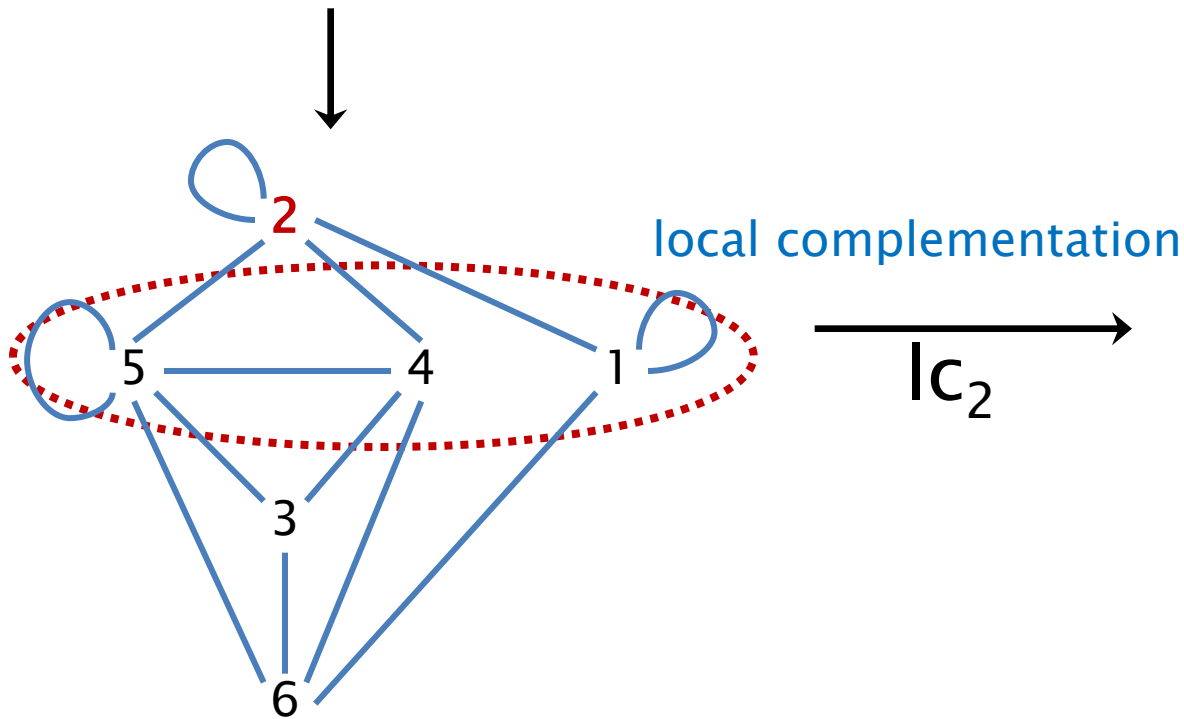


interlace information

# string to interlace graph

real generalization

$126\bar{1}34\bar{5}63\bar{2}45 \xrightarrow[\text{"invert"}]{\text{spr}_2} 12\bar{3}\bar{6}54\bar{3}1\bar{6}\bar{2}45$

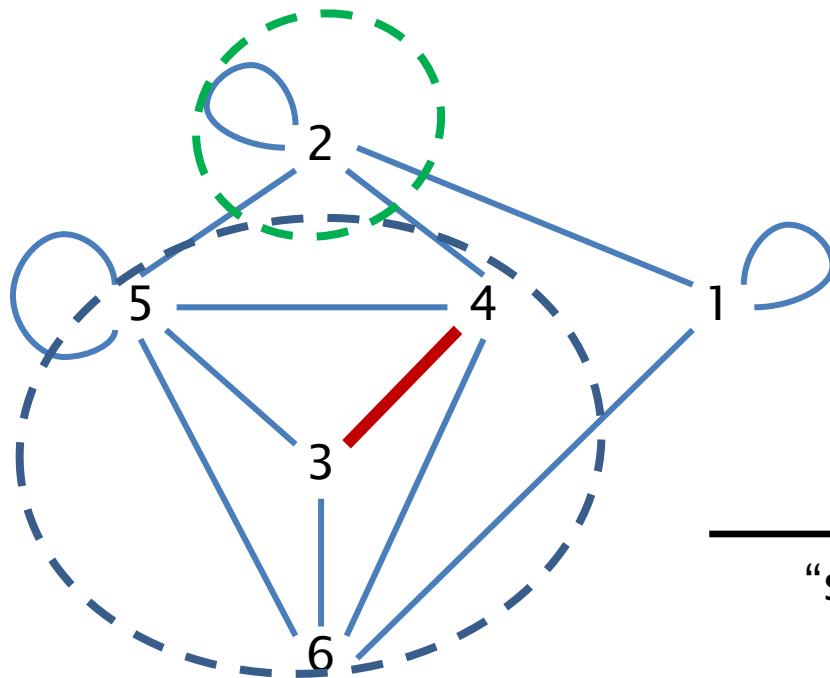
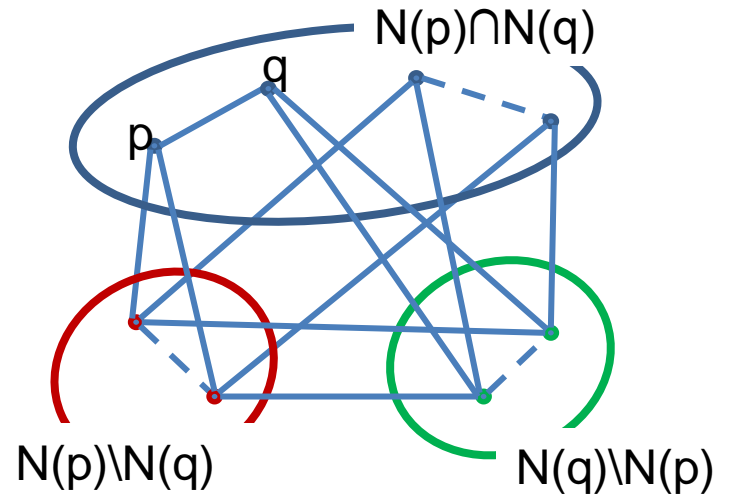


Kotzig 1966 without loops/bars

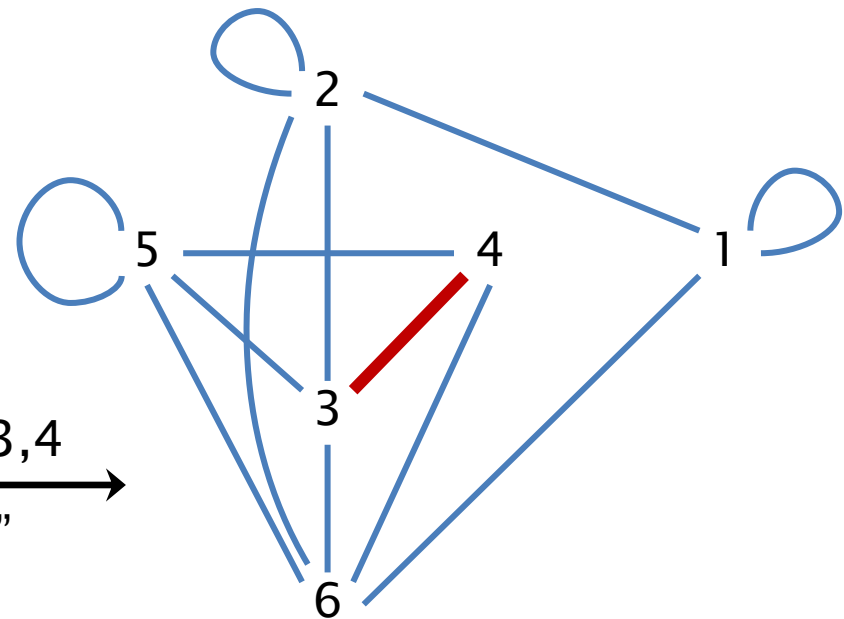
Ehrenfeucht et al, *Theor. Comp. Sci.*, 2003  
(for signed graphs instead of looped graphs)

# example edge local complementation

$\text{elc}_{3,4}$  on edge 3,4

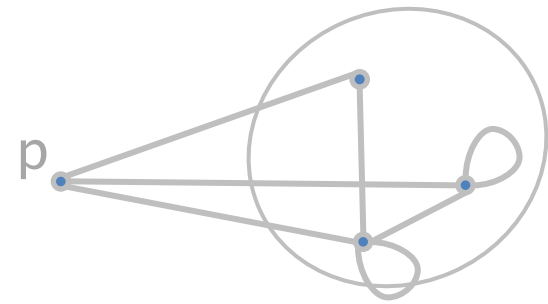
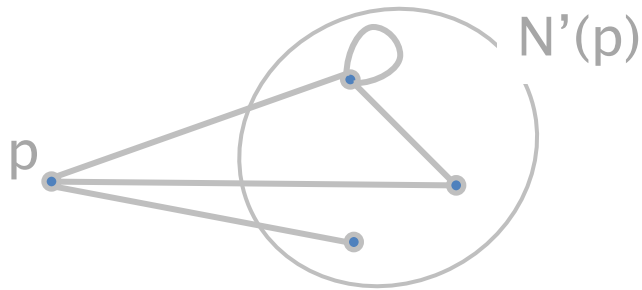


$\xrightarrow{\text{elc}_{3,4}}$   
"swap"

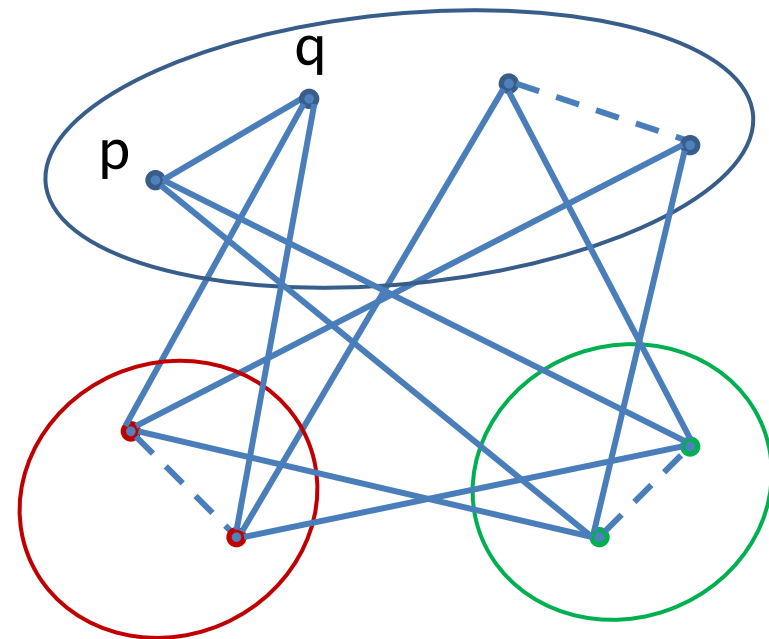
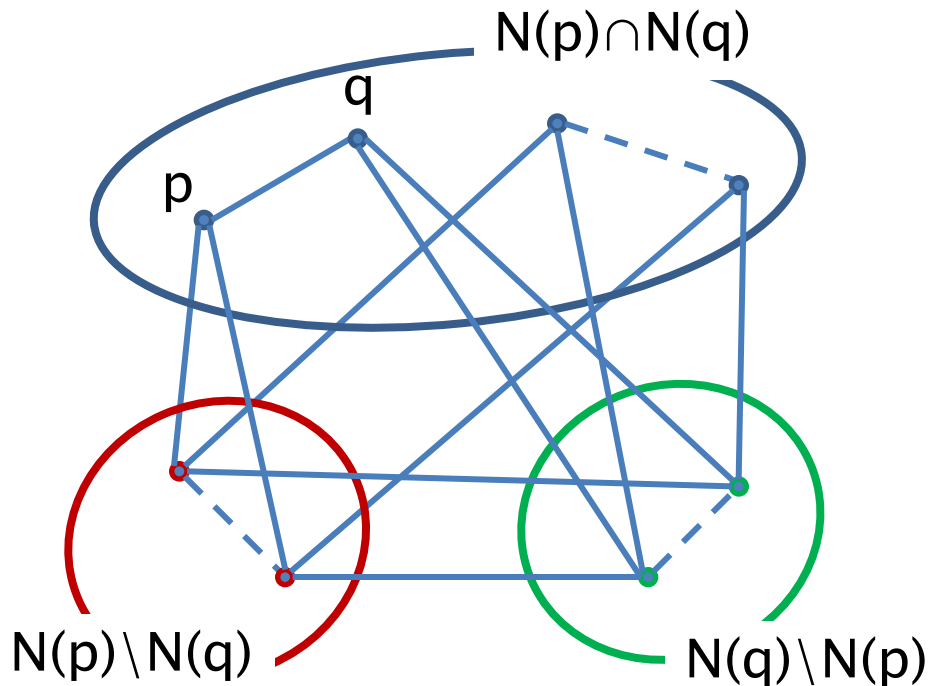


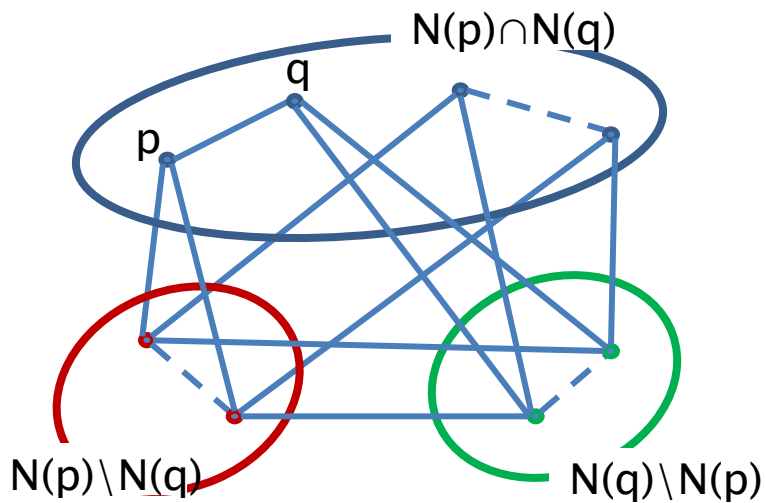
# graph operations

$lc_p$  *local complementation* looped vertex  $p$

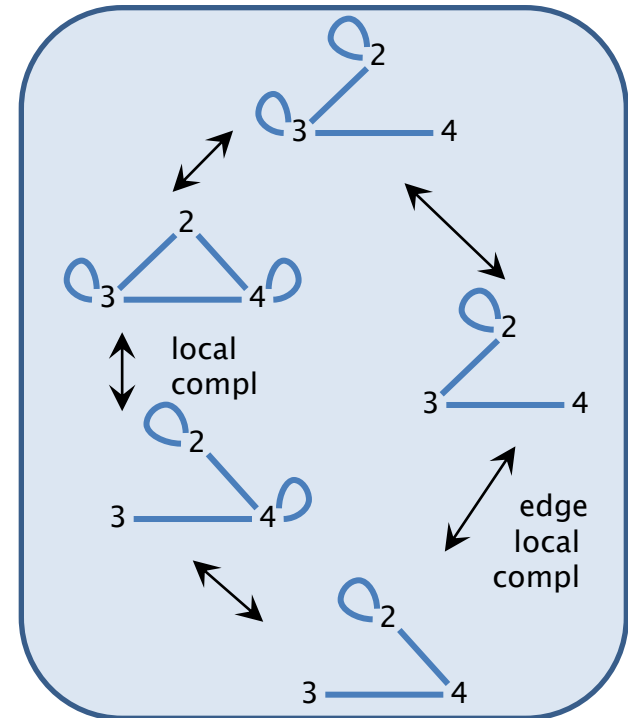


$elc_{p,q}$  *edge local complementation* unlooped edge  $pq$





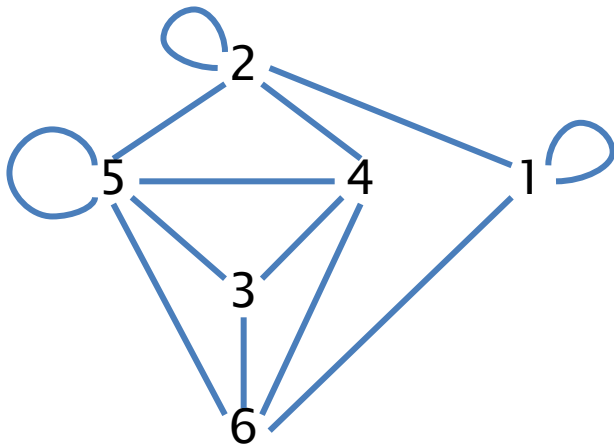
questions:



1) how do  $\text{elc}_{p,q}$  and  $\text{lc}_p$ , or  $\text{elc}_{p',q'}$  interact?

2) is the result of reductions dependent on (order) operations chosen? 4-regular graphs suggest: yes.

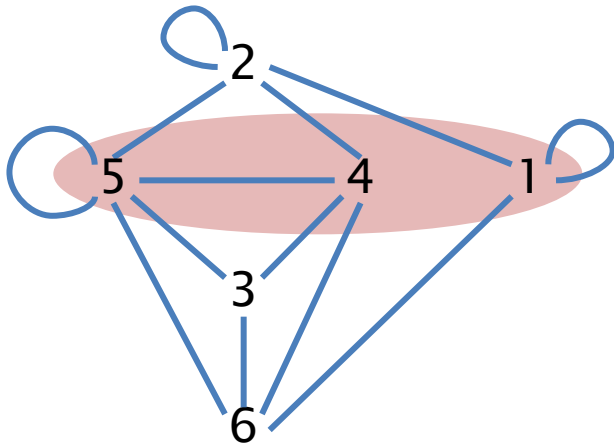
3) what are the intermediate products?



	1	2	3	4	5	6
1	1	1	0	0	0	1
2	1	1	0	1	1	0
3	0	0	0	1	1	1
4	0	1	1	0	1	1
5	0	1	1	1	1	1
6	1	0	1	1	1	0



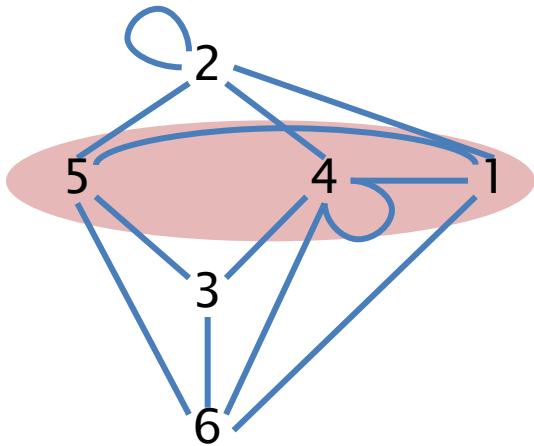
# reconsider local complementation



	1	2	3	4	5	6
1	1	1	0	0	0	1
2	1	1	0	1	1	0
3	0	0	0	1	1	1
4	0	1	1	0	1	1
5	0	1	1	1	1	1
6	1	0	1	1	1	0



$\downarrow$   $lc_2$



	1	2	3	4	5	6
1	0	1	0	1	1	1
2	1	1	0	1	1	0
3	0	0	0	1	1	1
4	1	1	1	1	0	1
5	1	1	1	0	0	1
6	1	0	1	1	1	0



$\downarrow$   $lc_2$

# (edge) local complementation is pivot

## principal pivot transform (ppt)

$A * X$  is defined iff  $A[X]$  is nonsingular

$$A = X \left( \begin{array}{c|c} & X \\ \hline P & Q \\ R & S \end{array} \right) \quad A * X = \left( \begin{array}{c|c} P^{-1} & -P^{-1} Q \\ \hline R P^{-1} & S - R P^{-1} Q \end{array} \right)$$

	1	2	3	4	5	6
1	1	1	0	0	0	1
2	1	1	0	1	1	0
3	0	0	0	1	1	1
4	0	1	1	0	1	1
5	0	1	1	1	1	1
6	1	0	1	1	1	0

$\xrightarrow{\text{elc}_{3,4}}$

	1	2	3	4	5	6
1	1	1	0	0	0	1
2	1	1	0	1	0	1
3	0	0	0	1	1	1
4	0	1	1	0	1	1
5	0	0	1	1	1	1
6	1	1	1	1	1	0

# principal pivot transform

partial inversion

$$A \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \text{ iff } A * X \begin{pmatrix} x_2 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_2 \end{pmatrix}$$

symmetric difference

$$(A * X) * Y = A * (X \Delta Y)$$

(when defined)

$$A * \{p_1, p_2\} \dots * \{p_n\} = A * V = A^{-1}$$

(all pointers)

any sequence involving all pointers

this shows that

- result does not depend on order of operations

# what *is* happening?

$$\begin{array}{c}
 342\bar{3}\bar{2}4 \\
 \begin{array}{c} 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 2 & 3 & 4 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}
 \end{array}
 \xrightarrow{lc_3 \ lc_4 \ lc_2}
 \begin{array}{c}
 3\bar{2}4234 \\
 \begin{array}{c} 2 \\ 3 \\ 4 \end{array} \begin{pmatrix} 2 & 3 & 4 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}
 \end{array}$$

multiply (over the binary numbers)

$$\begin{array}{c} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \text{micro} \end{array}
 \begin{array}{c} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \\ \text{macro} \end{array}
 =
 \begin{array}{c} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

+ xor  $\oplus$   $1+1=0$   
 \* and  $\wedge$

sorting DNA = computing the inverse

ppt is (partial) inverse

# four worlds

Micronuclear DNA

342 $\bar{3}$  $\bar{2}$ 4

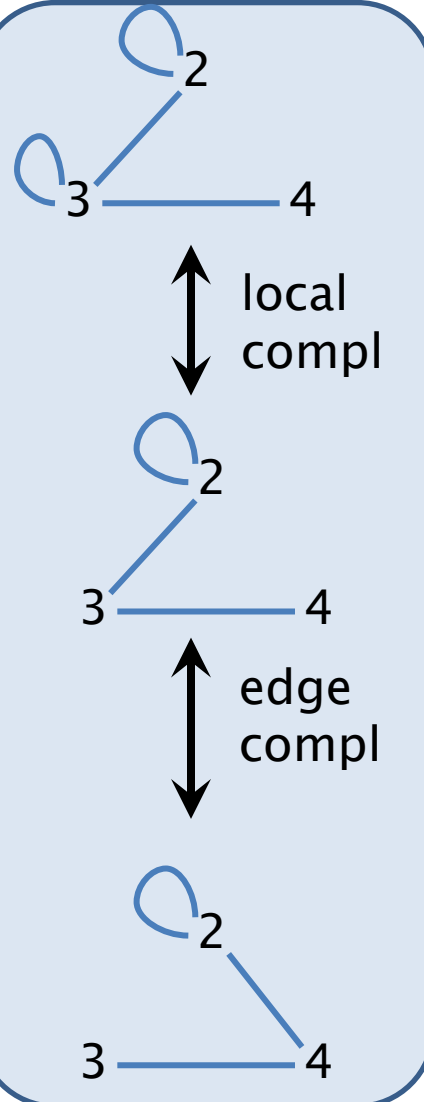
spr  $\updownarrow$  applicability restricted: enforce 1 string/intermolecular

3423 $\bar{2}$ 4

sdr  $\updownarrow$

3 $\bar{2}$ 4234

Macronuclear DNA



(interlace/circle) graphs

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

ppt  $\updownarrow$  \*{2}

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

\*{3,4}  $\updownarrow$

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

(binary) matrices

$\{\emptyset, \{2\}, \{3\}, \{3,4\}, \{2,3,4\}\}$

twist  $\updownarrow$   $\Delta\{2\}$

$\{\{2\}, \emptyset, \{2,3\}, \{2,3,4\}, \{3,4\}\}$

$\Delta\{3,4\}$   $\updownarrow$

$\{\{2,3,4\}, \{3,4\}, \{2,4\}, \{2\}, \emptyset\}$

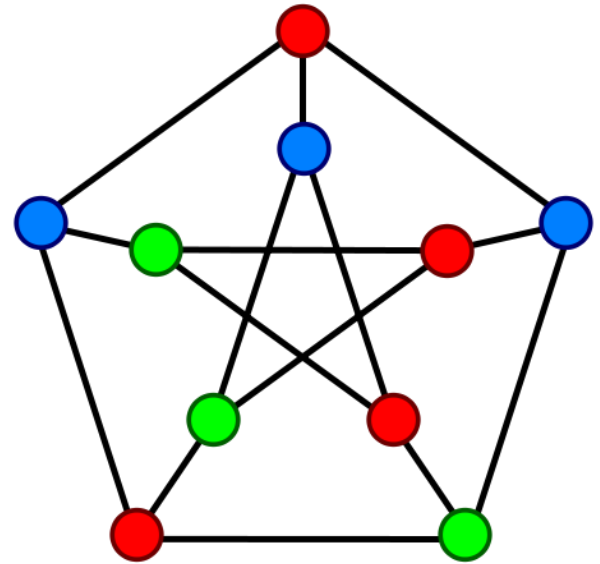
set systems (delta-)matroids

doc strings 4-regular graphs

- by careful modeling we find that gene assembly is *actually* principal pivot transform (ppt)
- we can use results about ppt to know more about gene assembly
  - independent order operations
  - interaction operations

# gene assembly polynomial

# chromatic polynomial



$$t(t-1)(t-2)(t^7 - 12t^6 + 67t^5 - 230t^4 + 529t^3 - 814t^2 + 775t - 352)$$

characteristic property

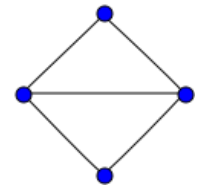
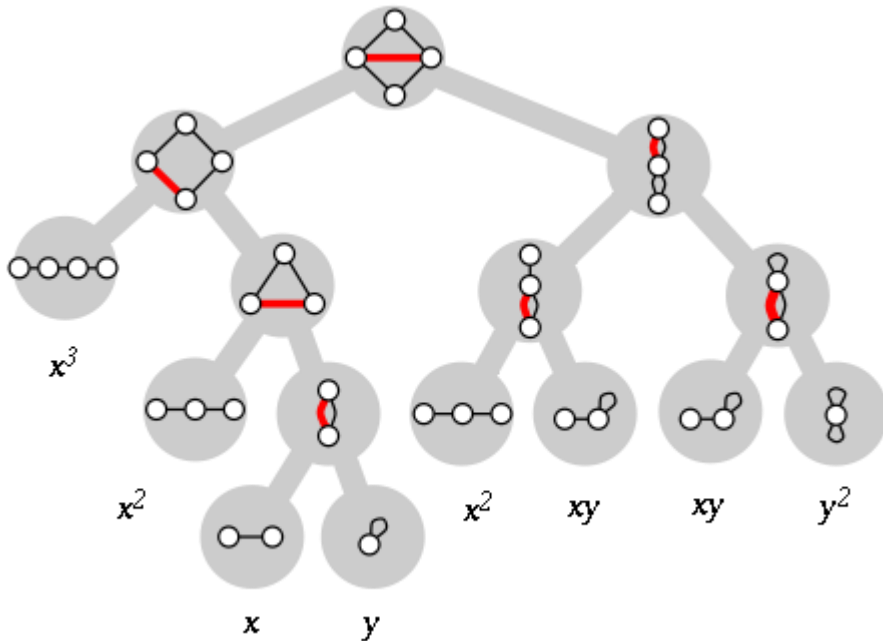


# Tutte polynomial

$$T_G(x, y) = \sum_{A \subseteq E} (x - 1)^{c(G[A]) - c(G)} (y - 1)^{c(G[A]) + |A| - |V|}$$

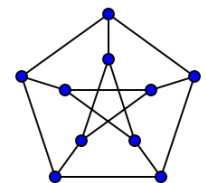
number components

of induced subgraph  $G[A]$



$$x^3 + 2x^2 + y^2 + 2xy + x + y$$

deletion-contraction algorithm

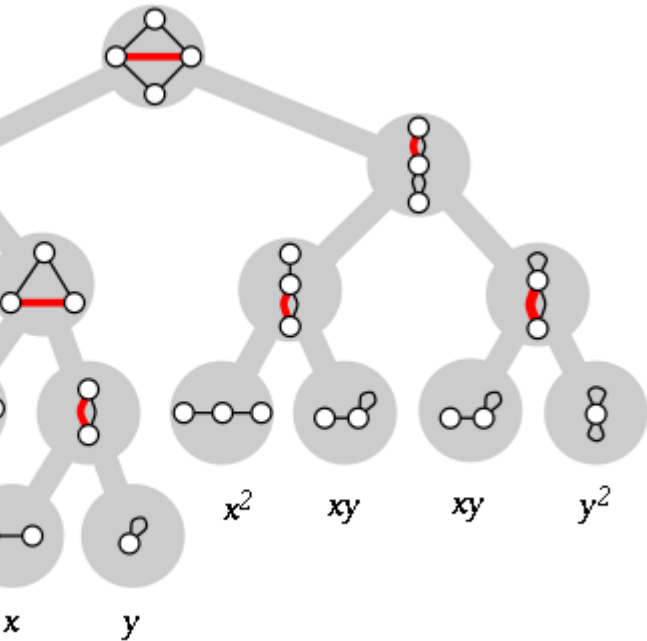


sage: P = graphs.PetersenGraph()

sage: P.tutte\_polynomial()

$x^9 + 6x^8 + 21x^7 + 56x^6 + 12x^5y + y^6 + 114x^5 + 70x^4y + 30x^3y^2 + 15x^2y^3 + 10xy^4 + 9y^5 + 170x^4 + 170x^3y + 105x^2y^2 + 65xy^3 + 35y^4 + 180x^3 + 240x^2y + 171xy^2 + 75y^3 + 120x^2 + 168xy + 84y^2 + 36x + 36y$

# graph polynomials



Tutte rather generic

- combinatorial
- algebraic
- recursive (deletion contraction)  
definitions / interpretations

interesting special values

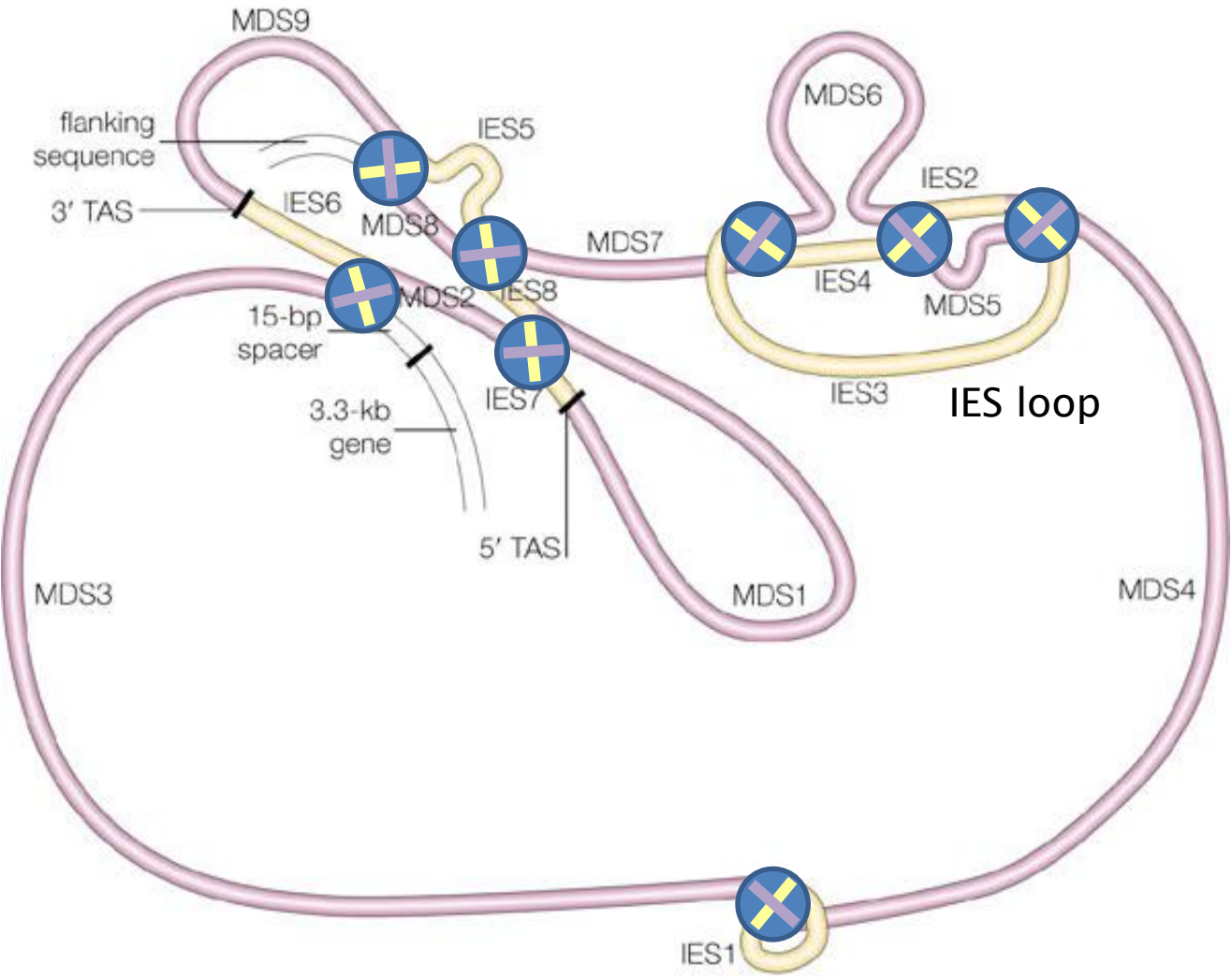
$T_G(1,1)$  # spanning forests

related

- knots “Jones polynomial”
- statistical physics

extended to matroids

# genome gymnastics






MAC

MDS and IES separated

⇒ motivation: counting intermediate products

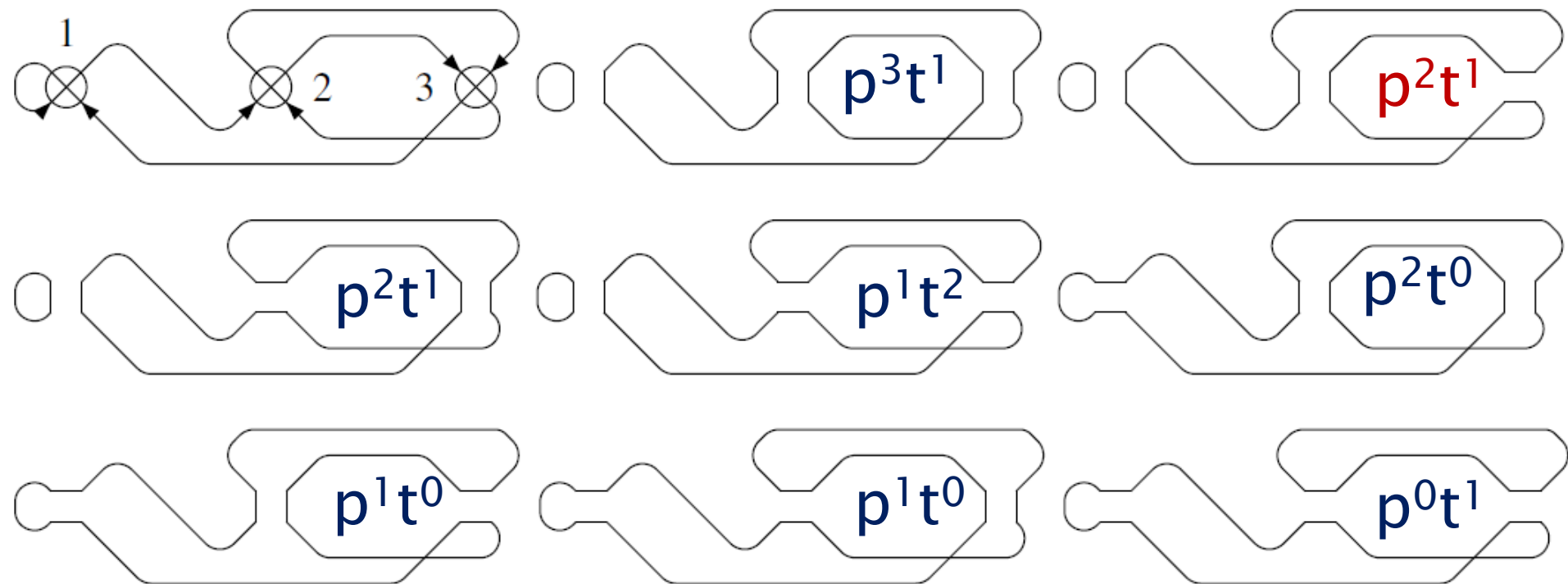
# assembly polynomial

$$w = 1 \ 1 \ 2 \ 3 \ 2 \ 3$$

-  follows C
-  consistent with C
-  inconsistent with C

$C_w$

2 consistent;  
2 components



$$p^3t + 2p^2t + p^2 + pt^2 + 2p + t$$

# assembly polynomial

doc string  $w$

$$S(w)(p, t) = \sum_s p^{\pi(s)} t^{c(s)-1}$$

summation over “transition systems”  
that never follow  $C_w$

$\pi(s)$  # orientation consistent trans.  
 $c(s)$  # components

# transition polynomials

F.Jaeger: On transition polynomials of 4-regular graphs (1990)

4-regular graph  
with *weighted* transitions

$$\sum_s \omega(s) t^{c(s)-1}$$

summation over *all* transition systems

- $\omega(s)$  product weights
- $c(s)$  # components

$$\begin{aligned}
 &= \text{circle with two loops} - \text{circle with two loops and a diagonal} - \text{circle with two loops and a diagonal} - \text{circle with two loops and a diagonal} + \text{circle with two loops and a diagonal} + \text{circle with two loops and a diagonal} + \text{circle with two loops and a diagonal} - \text{circle with two loops and a diagonal} \\
 &= 3^3 - 3^2 - 3^2 - 3^2 + 3 + 3 + 3 - 3
 \end{aligned}$$

# transition polynomials

4-regular graph

with *weighted* transitions

$$\sum_s \omega(s) t^{c(s)-1}$$

summation over *all* transition systems

-  $\omega(s)$  product weights

-  $c(s)$  # components

$$\sum_s p^{\pi(s)} t^{c(s)-1}$$

$\Rightarrow$  weights *assembly* pol:

0 follows C

p consistent with C

1 inconsistent with C

**corollary:**  $p=1$   $t=-2$

$$S(w)(1, -2) = (-1)^{|V|} (-2)^{c(P)-1} = (-1)^{|V|}$$

**Theorem (Traldi–Cohn–Lempel)**

- C Euler cycle in 4-regular graph
- P circuit partition, transitions wrt C
  - D<sub>1</sub> follows C
  - D<sub>2</sub> consistent with C
  - D<sub>3</sub> inconsistent with C

then

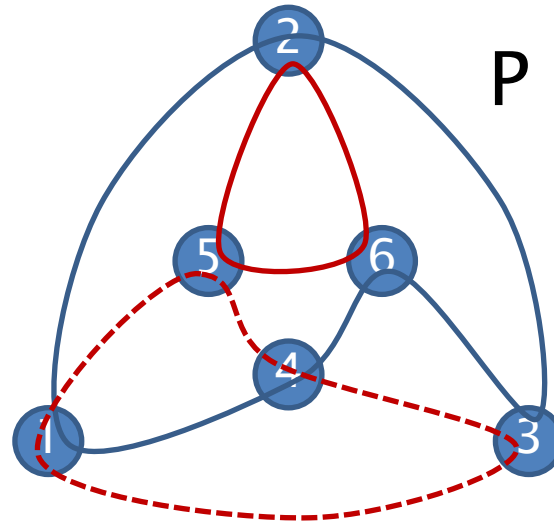
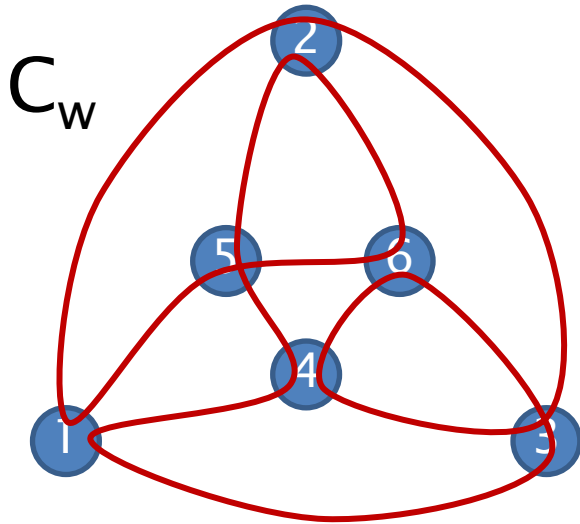
$$|P| = \nu \left( \underbrace{(I(C) + \Delta(D_3))}_{\text{overlap/interlace graph}} \setminus \underbrace{D_1}_{\text{deletion}} \right) + 1$$

loops /
diagonal matrix



# 4-regular graph from doc-string

$w = 1\ 4\ 5\ 2\ 6\ 5\ 1\ 2\ 3\ 4\ 6\ 3$



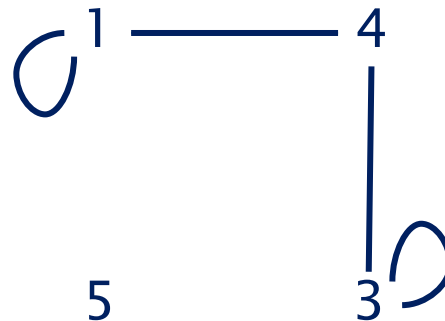
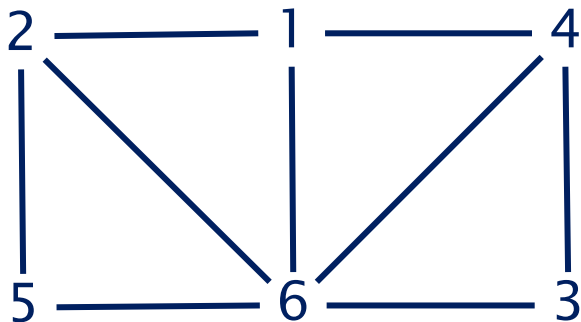
$$D_1 = \{ 2, 6 \}$$

$$D_2 = \{ 4, 5 \}$$

$$D_3 = \{ 1, 3 \}$$

$$|P| = \nu( (I(C) + \Delta(D_3)) \setminus D_1 ) + 1$$

$I(C_w)$



$$\begin{matrix}
 & 1 & 3 & 4 & 5 \\
 \begin{matrix} 1 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
 \end{matrix}$$

rank=2 nullity = 2

# algebraic formulation

$$S(w)(p, t) = \sum_s p^{\pi(s)} t^{c(s)-1}$$

“combinatorial”

- $D_1$  follows  $C$
- $D_2$  consistent with  $C$
- $D_3$  inconsistent with  $C$

$D_2$

$D_1 = \emptyset$   
 $D_3$

$$\sum_{D_2 \dot{\cup} D_3 = V} p^{|D_2|} t^{\nu(I(C_w) + \Delta(D_3))}$$

“algebraic”

what have we learned (for free)

## 1. as transition polynomial

- evaluations at special points
- recursive formulations  
“contraction–deletion”

## 2. algebraic formulation

using  $I(C_w)$  interlace graph

- related to bracket polynomial
- if  $I(C_{w_1})=I(C_{w_2})$  then equal pols

- by careful modeling we find that gene assembly is *actually* principal pivot transform (ppt)
- we can use results about ppt to know more about gene assembly
  - independent order operations
  - interaction operations
- assembly fit in a well-established field of graph polynomials

(end)



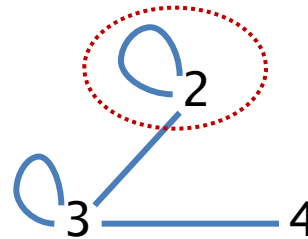
thanks!



$A * X$  is defined iff  $A[X]$  is nonsingular

$lc_2$

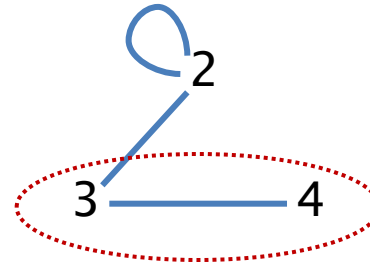
342 $\bar{3}$  $\bar{2}$ 4



$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

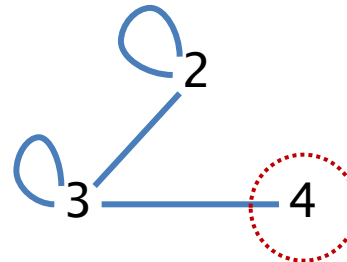
$elc_{3,4}$

3423 $\bar{2}$ 4



$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

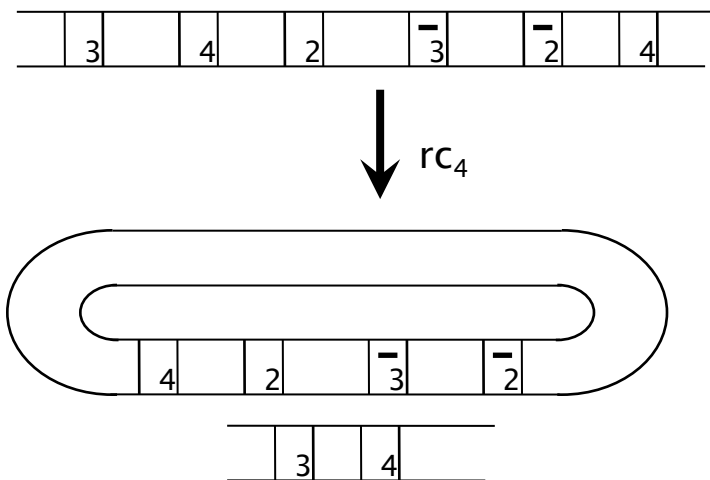
342 $\bar{3}$  $\bar{2}$ 4



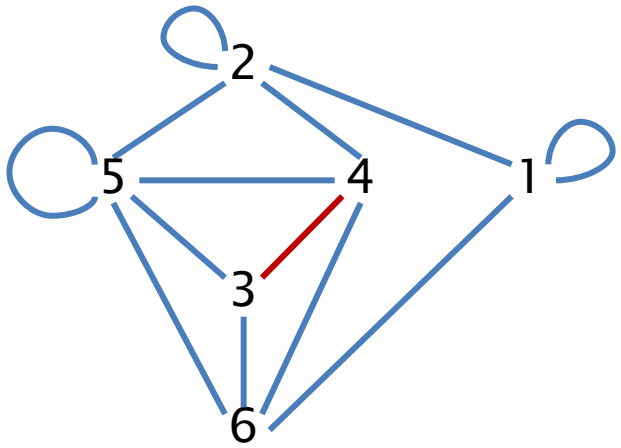
$$\begin{matrix} & \begin{matrix} 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

undefined

split not allowed



# reconsider edge local complementation

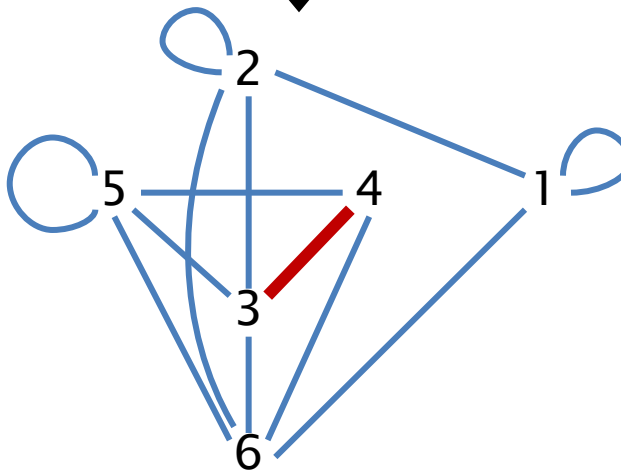


	1	2	3	4	5	6
1	1	1	0	0	0	1
2	1	1	0	1	1	0
3	0	0	0	1	1	1
4	0	1	1	0	1	1
5	0	1	1	1	1	1
6	1	0	1	1	1	0



elc<sub>3,4</sub>

elc<sub>3,4</sub>



	1	2	3	4	5	6
1	1	1	0	0	0	1
2	1	1	0	1	0	1
3	0	0	0	1	1	1
4	0	1	1	0	1	1
5	0	0	1	1	1	1
6	1	1	1	1	1	0





$$t(t-1)(t-2)(t^7-12t^6+67t^5-230t^4+529t^3-814t^2+775t-352)$$

$$|P| = \nu(\Delta(D_3) \setminus D_1) + 1$$

<http://www.codecogs.com/latex/eqneditor.php>

computer modern

Tutte polynomial

$$x^3+2x^2+y^2+2xy+x+y$$

@(1,1) spanning forests

$$T_G(x,y) = \sum_{A \subseteq E} (x-1)^{k(A)-k(E)} (y-1)^{k(A)+|A|-|V|}$$

$$S(G_w)(p,t) = \sum_s p^{\pi(s)} t^{c(s)}$$

$$S(w)(p,t) = \sum_s p^{\pi(s)} t^{c(s)-1}$$

$$p^3t+2p^2t+p^2+pt^2+2p+t$$

application  $p=1$   $t=-2$       $S(w)(1,-2) = (-1)^{|V|} (-2)^{c(P)-1} = (-1)^{|V|}$

$$\sum_s \omega(s) t^{c(s)-1}$$

### this presentation

R. Brijder, H.J. Hoogeboom. The Algebra of Gene Assembly in Ciliates. In: *Discrete and Topological Models in Molecular Biology* (Jonoska and Saito, eds.) Springer (to appear).

### recombination is ppt

R. Brijder, T. Harju, H.J. Hoogeboom, Pivots, determinants, and perfect matchings of graphs (2008) [arXiv:0811.3500] TCS (2012).

<http://dx.doi.org/10.1016/j.tcs.2012.02.031>

(independently) Nathan Pflueger: Graph reductions, binary rank, and pivots in gene assembly. DAM 159 (2011) 2117–2134.

<http://dx.doi.org/--->

### general

R. Brijder, M. Daley, T. Harju, N. Jonoska, I. Petre, G. Rozenberg: Computational nature of gene assembly in ciliates. Chapter 37 in *Handbook of Natural Computing* (Rozenberg, Bäck, Kok, eds, Springer (2012)).

A. Ehrenfeucht, T. Harju, I. Petre, D. Prescott, G. Rozenberg, *Computation in Living Cells: Gene Assembly in Ciliates*, Natural Computing Series, Springer (2004).