

Computability in Europe 2014

(session on Bio-inspired computation)

Budapest, Hungary

June 2014

DLF Lunch-Lezing
1 oktober 2014

Graph polynomials motivated by Gene Rearrangements in Ciliates

Hendrik Jan Hoogeboom

Leiden NL

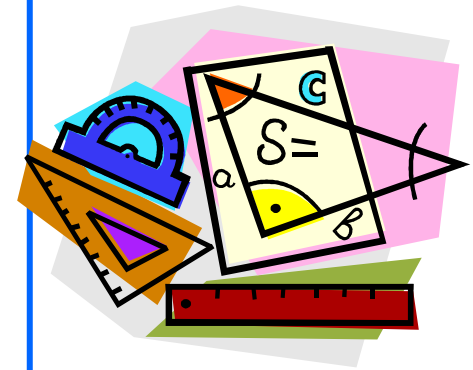
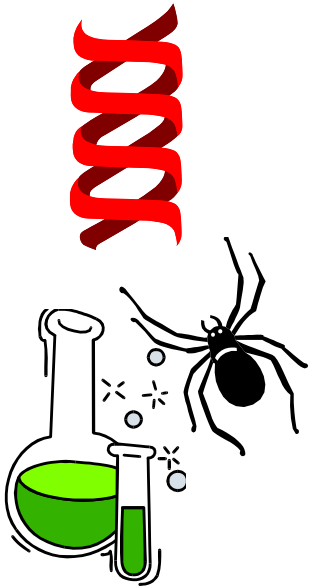
joint with

Robert Brijder

Hasselt B

natural computation

bio inspired computing



comput molecular biol

DNA computing

DNA en grafen

inhoudsopgave



↑
pizza canibale

inhoudsopgave

unscrambling genes
in ciliates

de Bruijn graphs &
gene assembly

sorting by reversal

inhoudsopgave

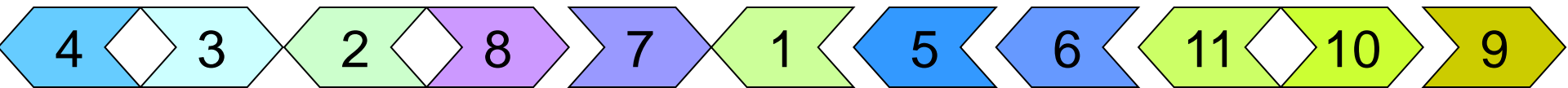
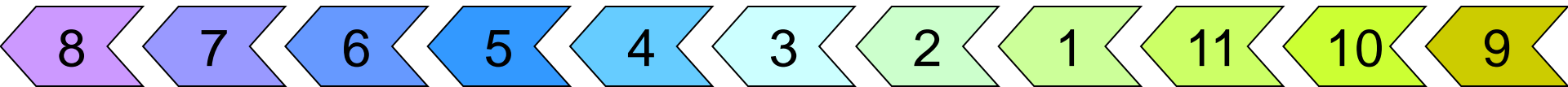
unscrambling genes
in ciliates

de Bruijn graphs &
gene assembly

sorting by reversal

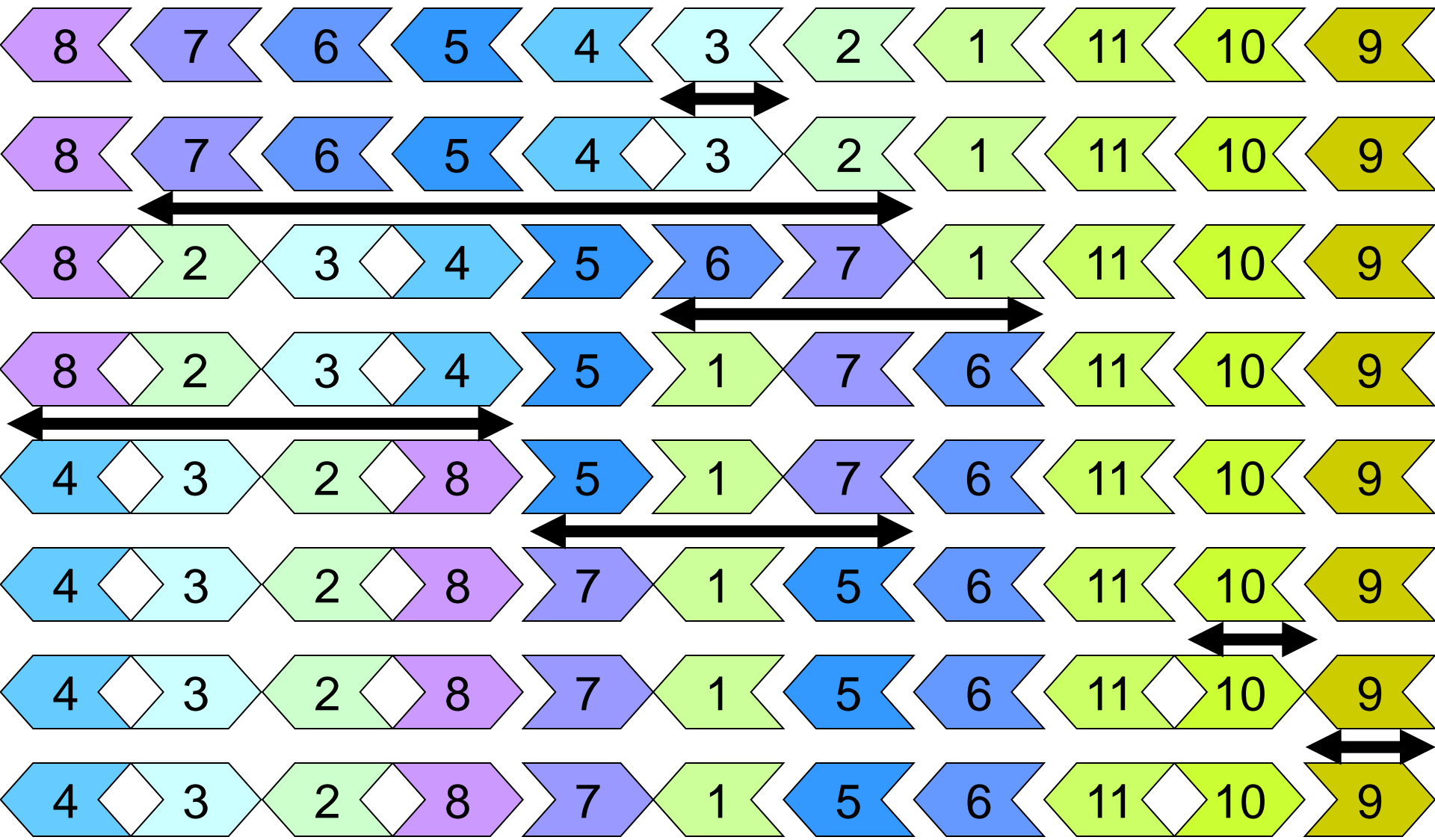
genome: turning cabbage into turnip

'van kool naar raap'



AGAGGATCCTTGCTGGATCCTGA
TCTCCTAGGAACGACC TAGGACT

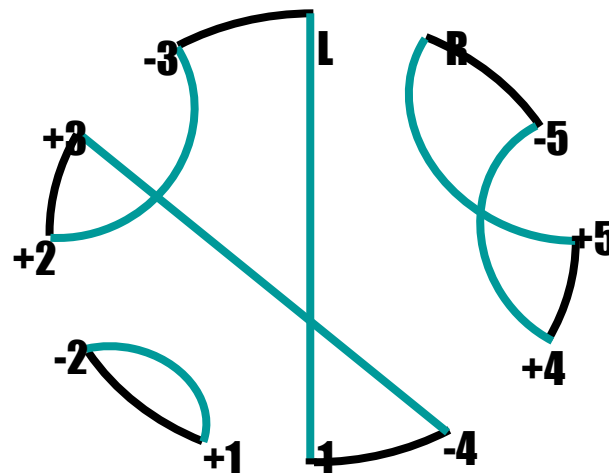
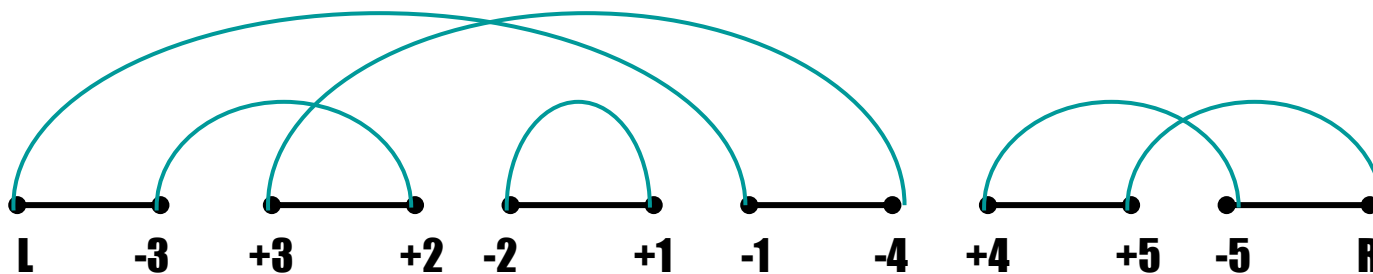
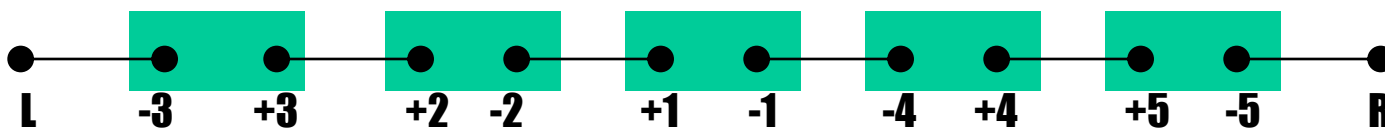
genome: turning cabbage into turnip



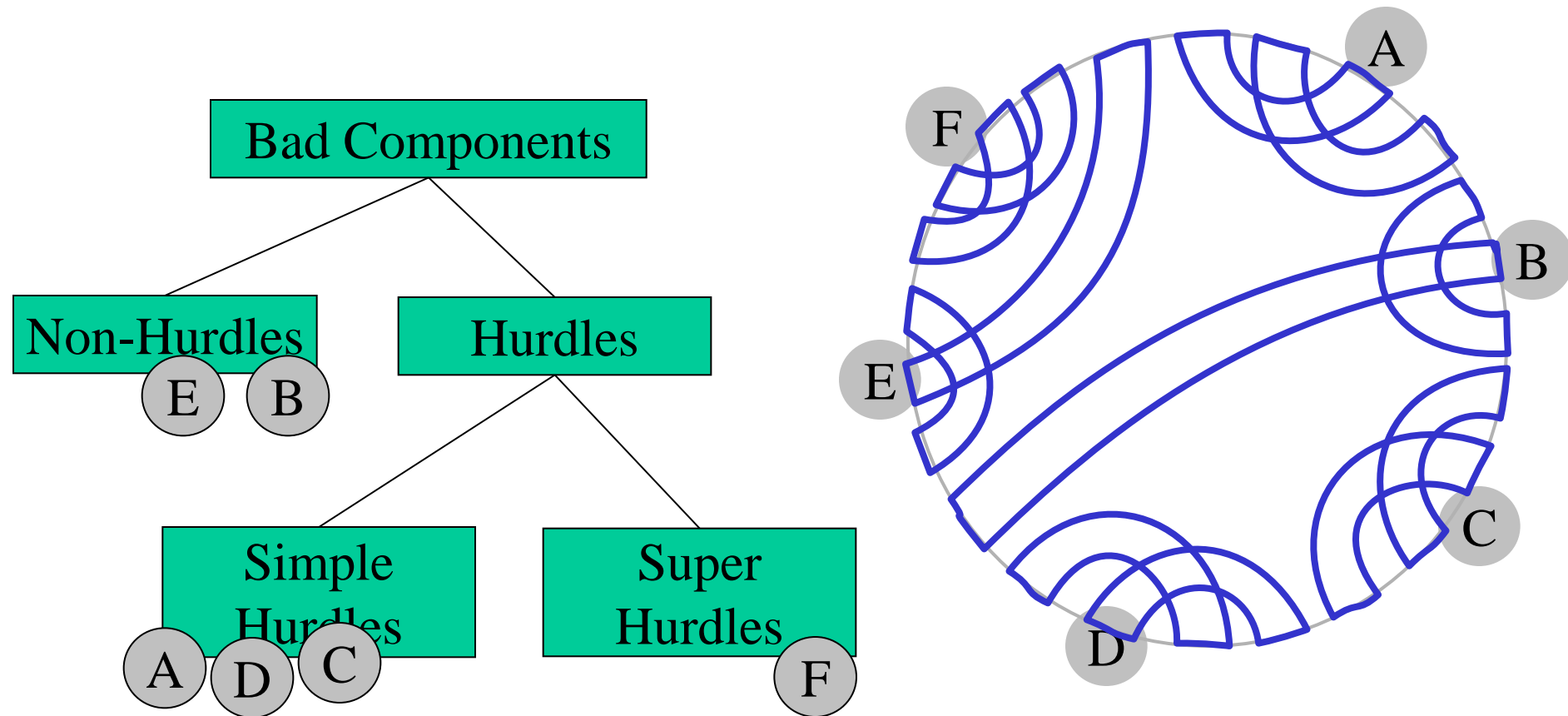
mitochondrial DNA: worm Ascaris Suum into human

12 31 34 28 26 17 29 4 9 36 18 35 19 1 16 14 32 33 22 15 11 27 5 20 13 30 23 10 6 3 24 21 8 25 2 7
12 31 34 28 26 17 29 4 9 36 18 35 19 1 16 14 33 32 22 15 11 27 5 20 13 30 23 10 6 3 24 21 8 25 2 7
12 31 32 33 14 16 1 19 35 18 36 9 4 29 17 26 28 34 22 15 11 27 5 20 13 30 23 10 6 3 24 21 8 25 2 7
12 33 32 31 30 13 20 5 27 11 15 22 34 28 26 17 29 4 9 36 18 35 19 1 16 14 23 10 6 3 24 21 8 25 2 7
12 33 32 31 30 29 17 26 28 34 22 15 11 27 5 20 13 4 9 36 18 35 19 1 16 14 23 10 6 3 24 21 8 25 2 7
12 33 32 31 30 29 28 26 17 34 22 15 11 27 5 20 13 4 9 36 18 35 19 1 16 14 23 10 6 3 24 21 8 25 2 7
12 33 32 31 30 29 28 27 11 15 22 34 17 26 5 20 13 4 9 36 18 35 19 1 16 14 23 10 6 3 24 21 8 25 2 7
12 33 32 31 30 29 28 27 26 17 34 22 15 11 5 20 13 4 9 36 18 35 19 1 16 14 23 10 6 3 24 21 8 25 2 7
12 33 32 31 30 29 28 27 26 25 8 21 24 3 6 10 23 14 16 1 19 35 18 36 9 4 13 20 5 11 15 22 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 21 8 3 6 10 23 14 16 1 19 35 18 36 9 4 13 20 5 11 15 22 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 23 10 6 3 8 21 14 16 1 19 35 18 36 9 4 13 20 5 11 15 22 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 15 11 5 20 13 4 9 36 18 35 19 1 16 14 21 8 3 6 10 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 14 16 1 19 35 18 36 9 4 13 20 5 11 15 8 3 6 10 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 13 4 9 36 18 35 19 1 16 14 5 11 15 8 3 6 10 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 35 18 36 9 4 13 1 16 14 5 11 15 8 3 6 10 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 35 36 9 4 13 1 16 14 5 11 15 8 3 6 10 34 17 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 34 10 6 3 8 15 11 5 14 16 1 13 4 9 36 35 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 14 5 11 15 8 3 6 10 34 1 13 4 9 36 35 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 11 5 14 8 3 6 10 34 1 13 4 9 36 35 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 5 11 8 3 6 10 34 1 13 4 9 36 35 2 7
12 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 1 34 10 6 3 8 11 5 4 9 36 35 2 7
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 1 34 10 6 3 4 5 11 8 9 36 35 2 7
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 1 34 35 36 9 8 11 5 4 3 6 10 2 7
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 1 34 35 36 9 8 7 2 10 6 3 4 5 11
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 1 34 35 36 9 8 7 6 10 2 3 4 5 11
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 1 34 35 36 9 8 7 6 5 4 3 2 10 11
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 1 2 3 4 5 6 7 8 9 36 35 34 10 11
12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 9 8 7 6 5 4 3 2 1 10 11
1 2 3 4 5 6 7 8 9 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36

model: *reality and desire*

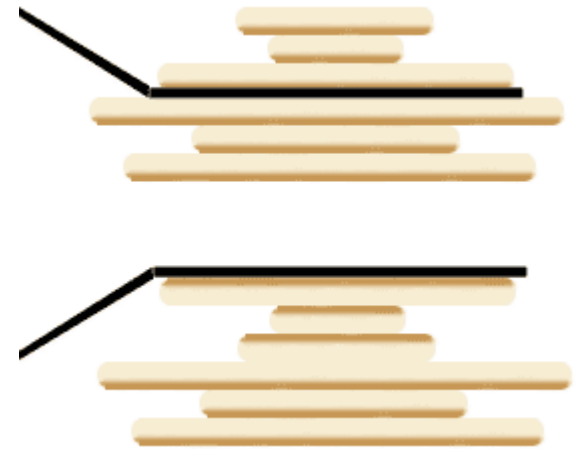


hurdle & fortress



$$d(\pi) \geq b(\pi) - c(\pi) + h(\pi) + f(\pi)$$

pancake sort



Gates, W., Papadimitriou, C. (1979).
"Bounds for Sorting by Prefix Reversal".
Discrete Mathematics 27: 47–57.

inhoudsopgave

unscrambling genes
in ciliates

de Bruijn graphs &
gene assembly

sorting by reversal

Nicolaas Govert "Dick" de Bruijn

(9 July 1918 – 17 February 2012)

- De Bruijn sequence (1946)
- BEST Theorem (1951)
- Automath (1967)



<http://www.win.tue.nl/automath/images/photos/deBruijn.jpg>

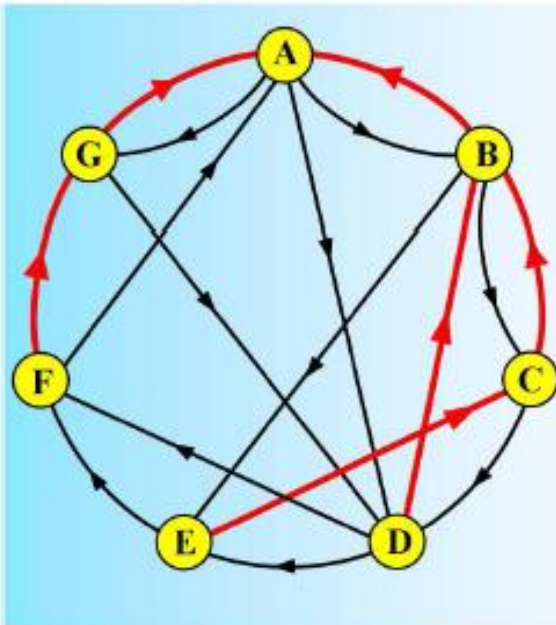
de Bruijn, van Aardenne-Ehrenfest, Smith & Tutte

<http://spikedmath.com/327.html>

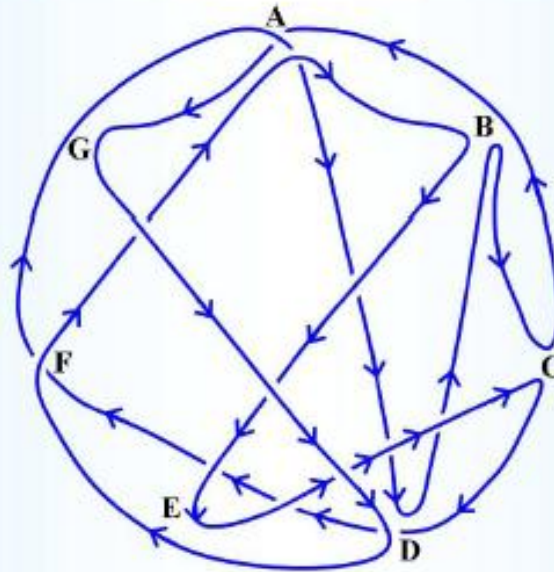
$$ec(G) = t_w(G) \prod_{v \in V} (\deg(v) - 1)!$$

spikedmath.com
© 2010

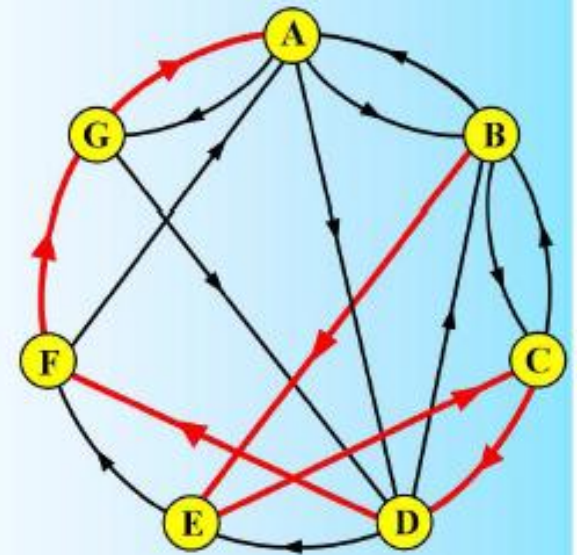
**BEST.
THEOREM.
EVER!**



directed trees



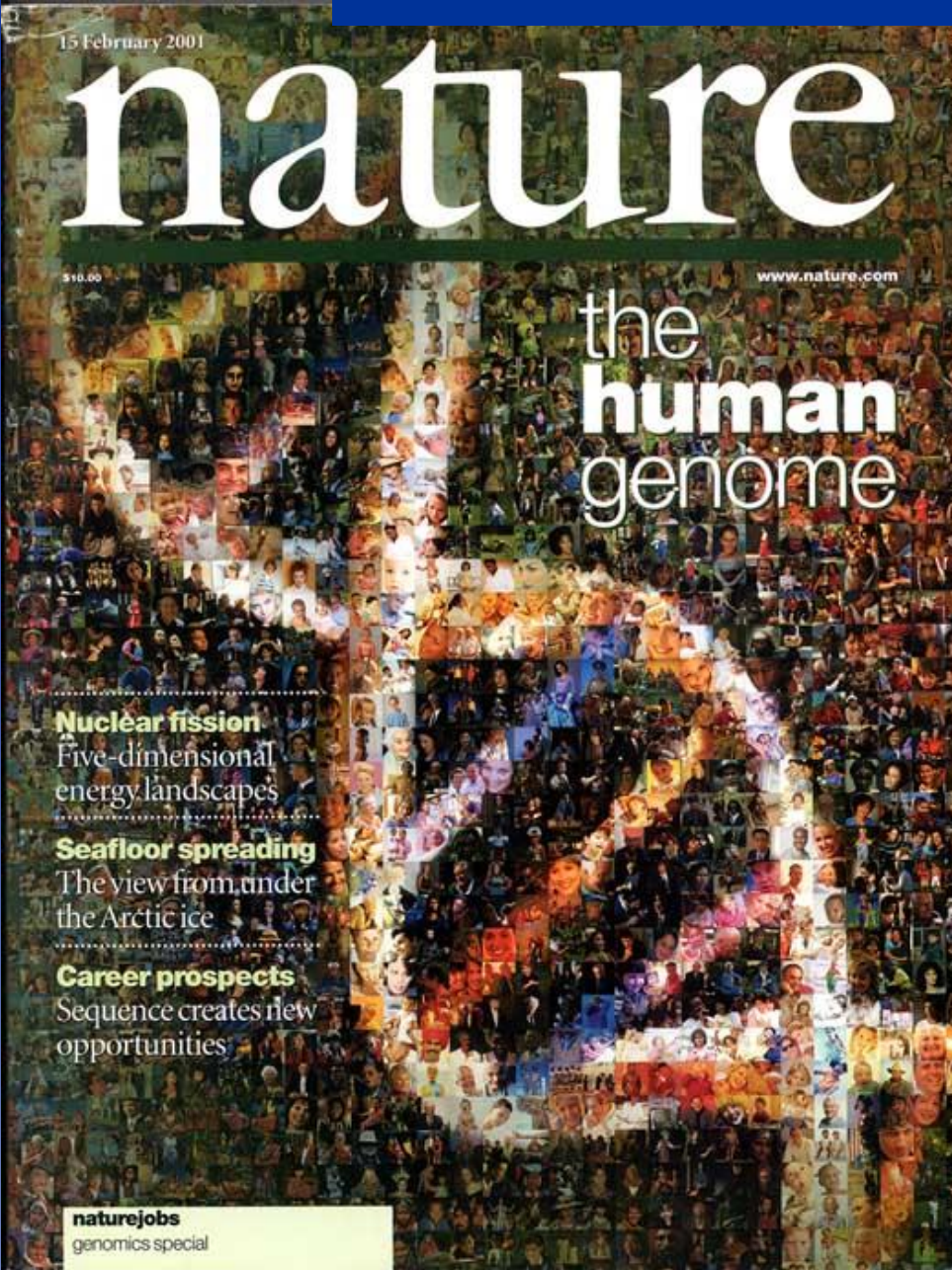
euler cycles



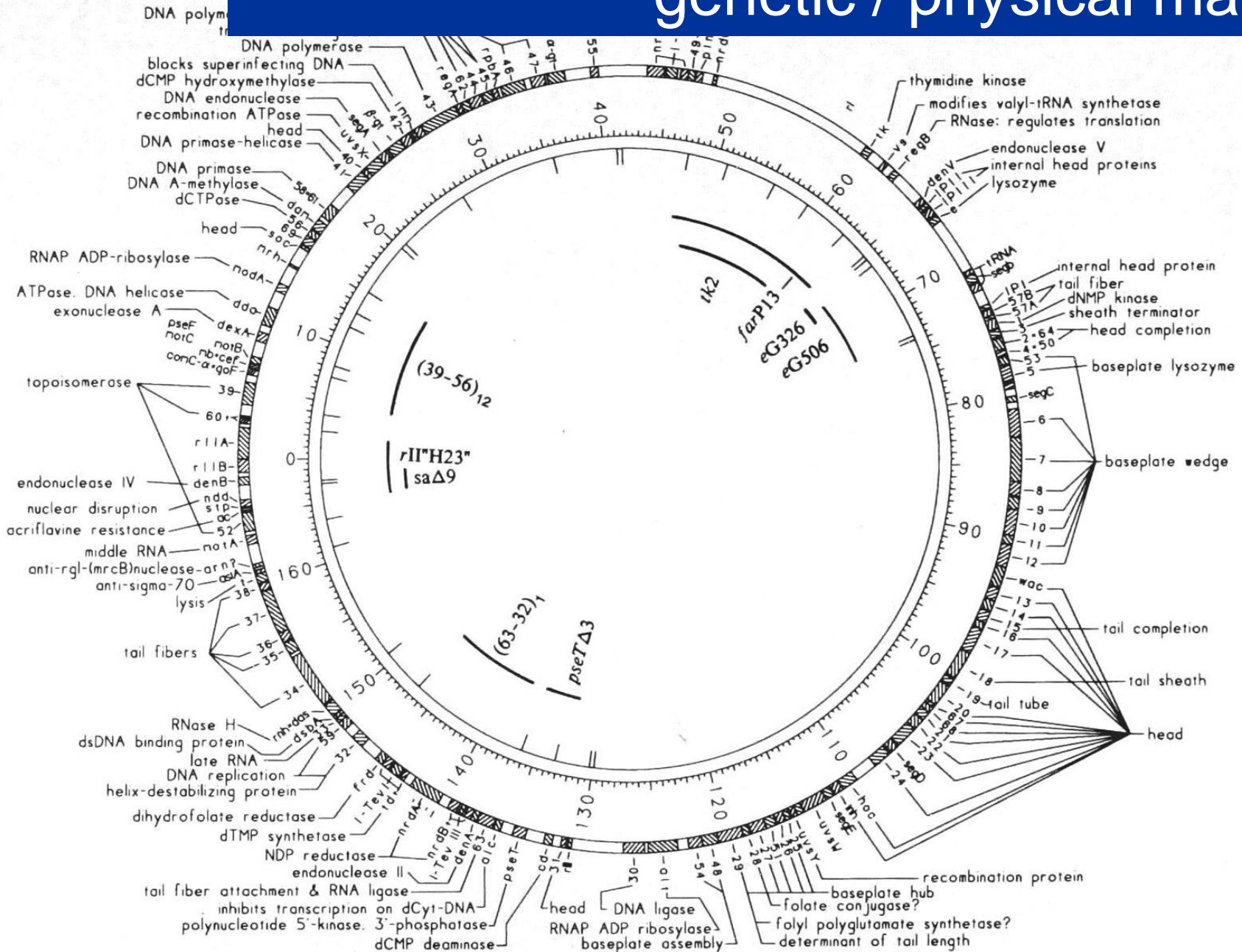
speciaal geval $\deg(v)=2$

<http://myweb.lsbu.ac.uk/~whitty/MathSci/TheoremOfTheDay/CombinatorialTheory/BEST/TotDBEST.pdf>

feb'01 - human genome



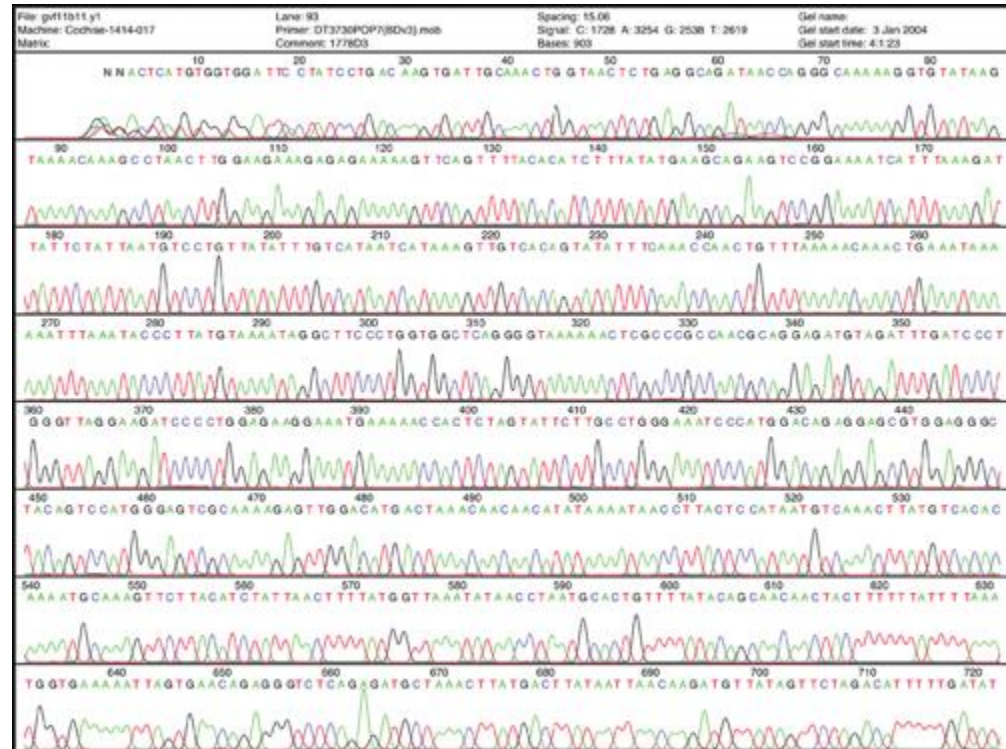
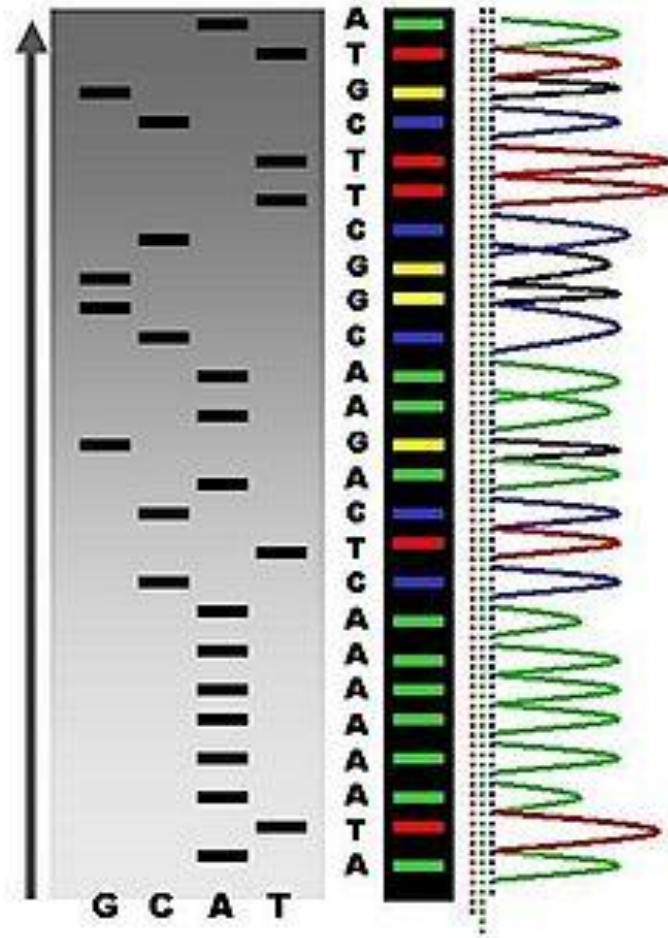
genetic / physical map

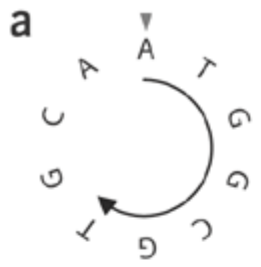


chain-termination sequencing

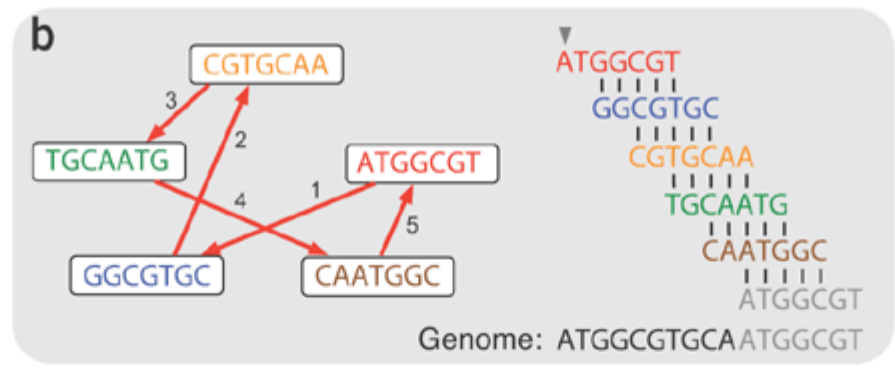
(Sanger)

300 – 1000 mer



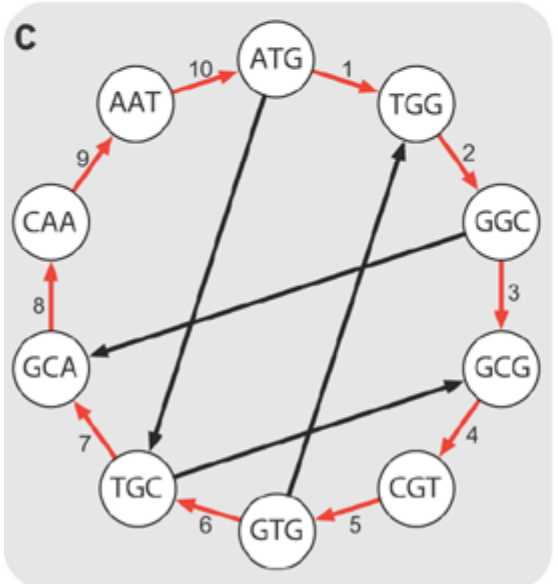


Short-read sequencing

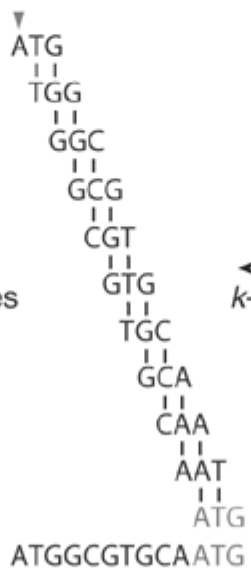


Vertices are k -mers
Edges are pairwise alignments

Vertices are $(k-1)$ -mers
Edges are k -mers

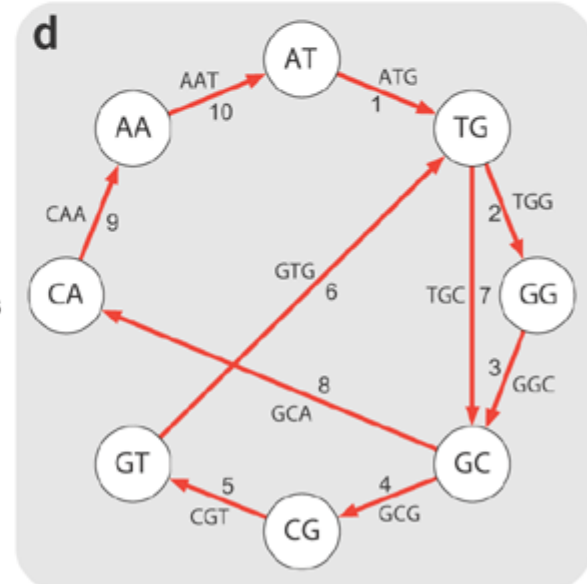


Hamiltonian cycle
Visit each vertex once
(harder to solve)



k -mers from vertices

k -mers from edges



Eulerian cycle
Visit each edge once
(easier to solve)

How to apply de Bruijn graphs to genome assembly
(Phillip E C Compeau, Pavel A Pevzner & Glenn Tesler)

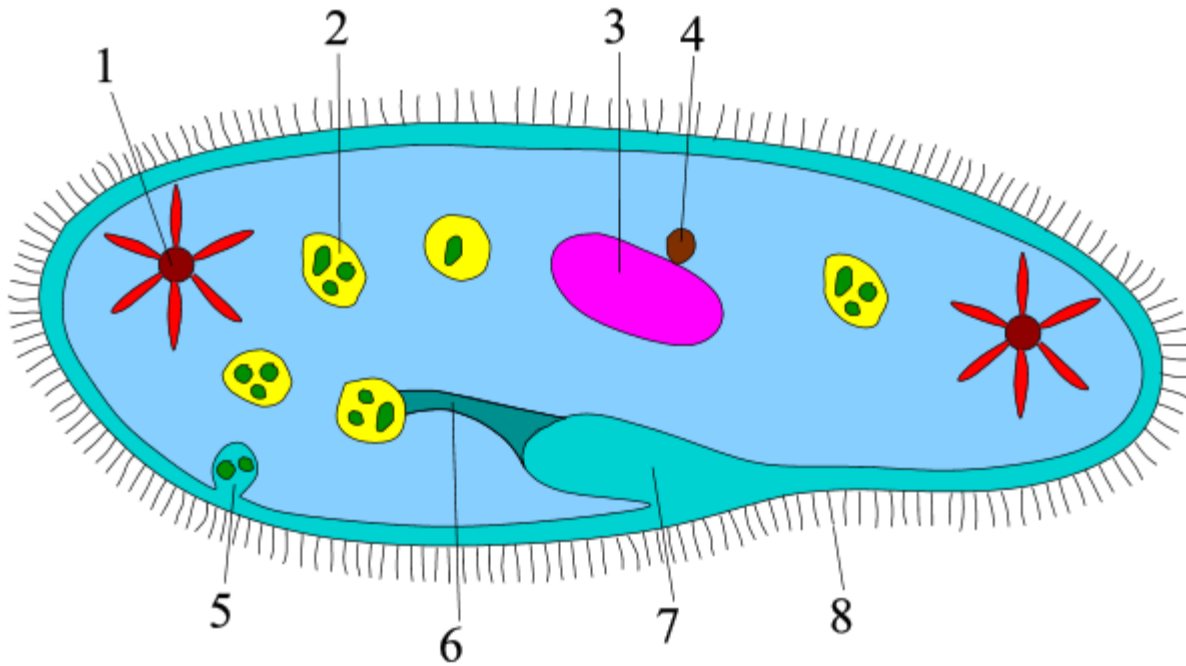
inhoudsopgave

unscrambling genes
in ciliates

de Bruijn graphs &
gene assembly

sorting by reversal

ciliates: micro and macro



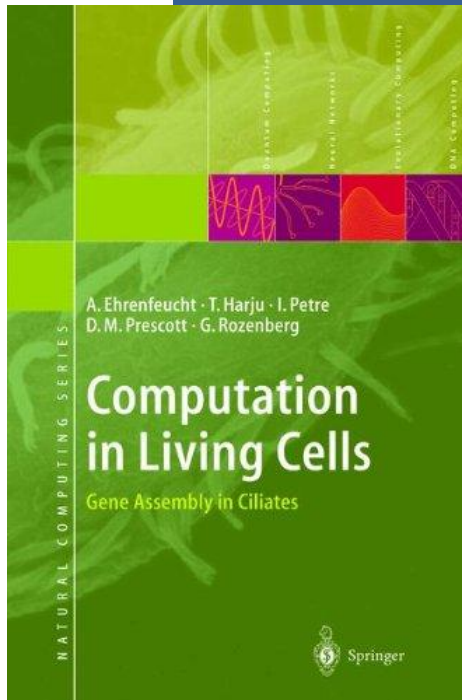
cell structure:

- 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



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”

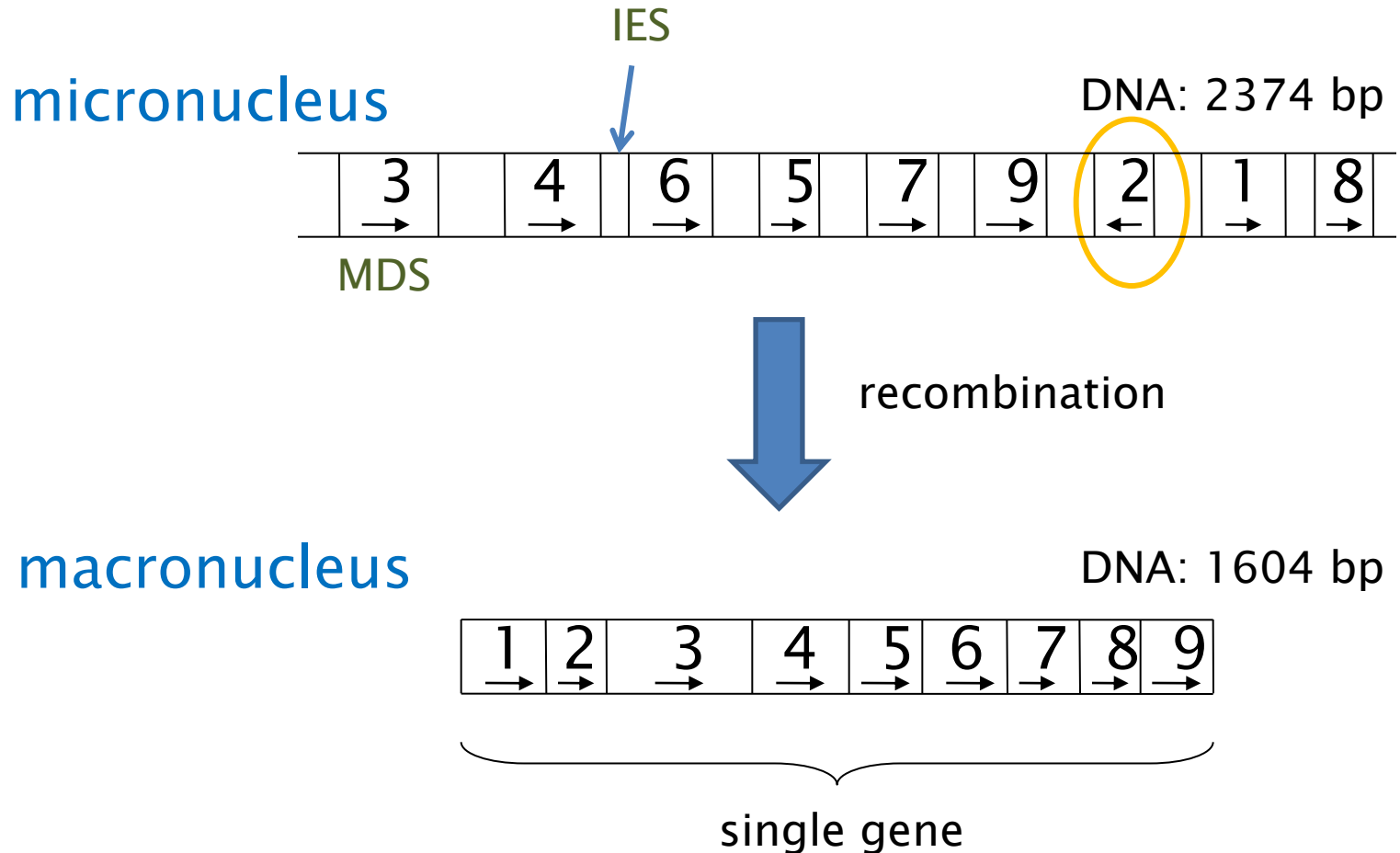
[L. Landweber, L. Kari: Universal Molecular Computation in Ciliates, 2002]

“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]

from micro to macro

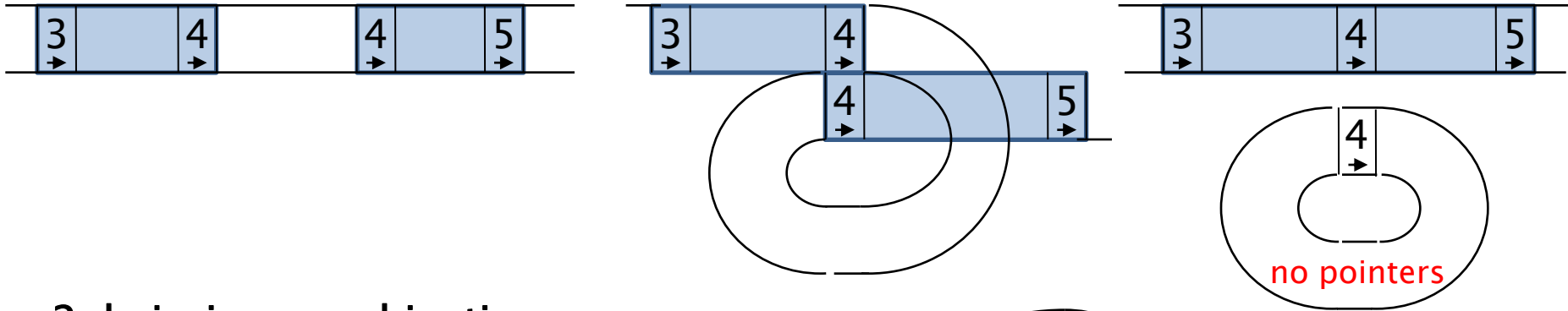
http://oxytricha.princeton.edu/cgi-bin/get_MDS_IES_Info.cgi?num=38



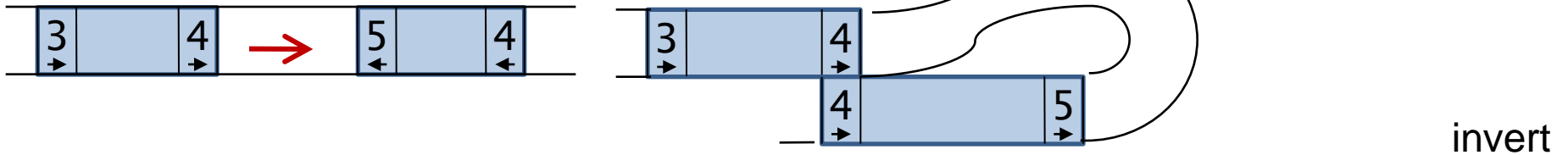
Greslin, Prescott et al. Reordering of nine exons is necessary to form a functional actin gene in *Oxytricha nova*. PNAS 86, 6264–6268, Aug 1989.

recombination on pointers

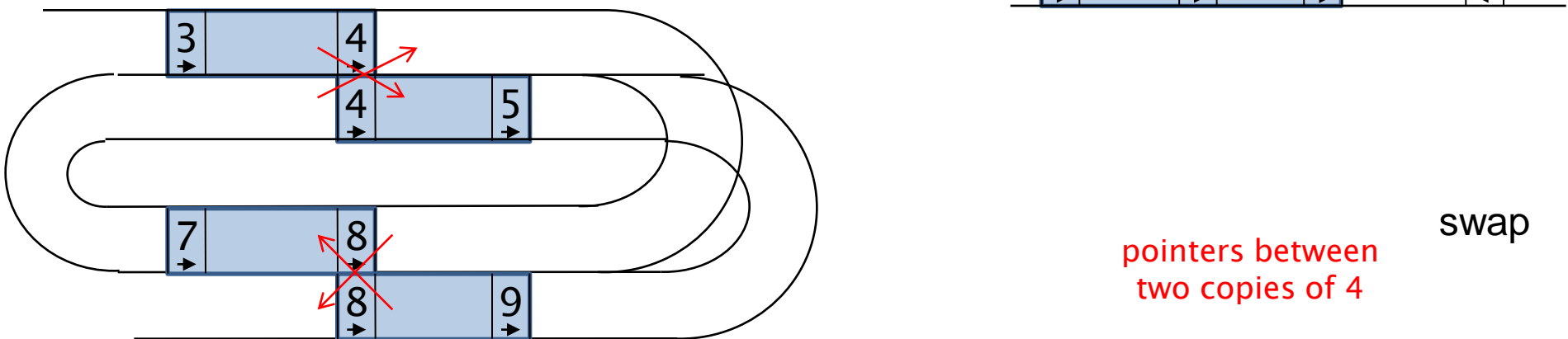
1. loop recombination



2. hairpin recombination



3. double-loop recombination

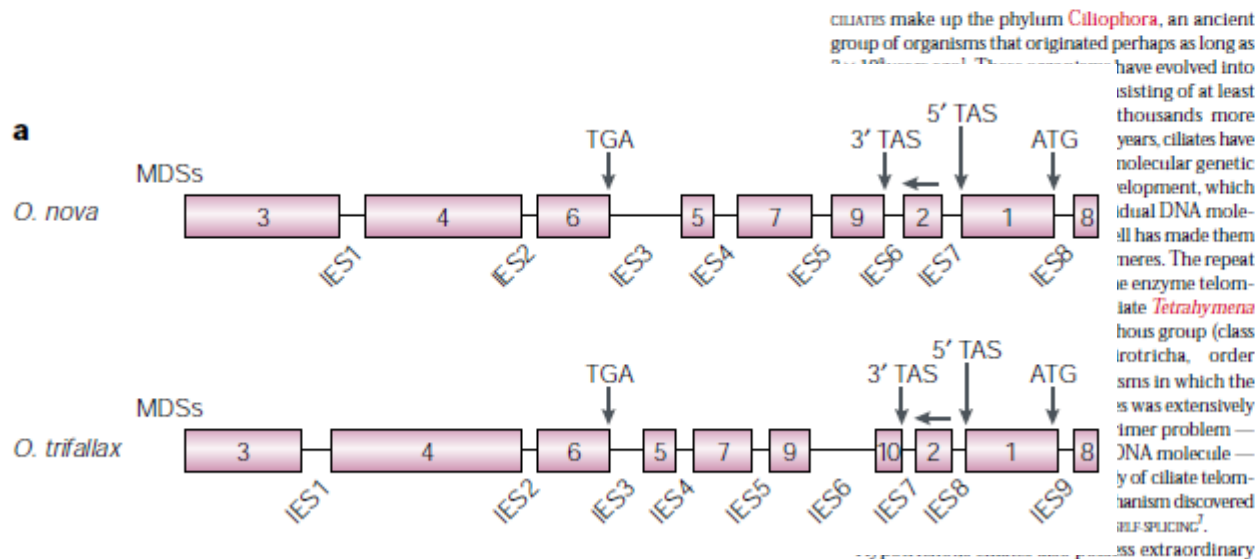


look at the “gymnastics”

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 rDNA, order trichostema, order trichostema) is extensively used in the trichostema group (class I rDNA, order trichostema) of ciliate telomerase discovered in 1987. This extraordinary

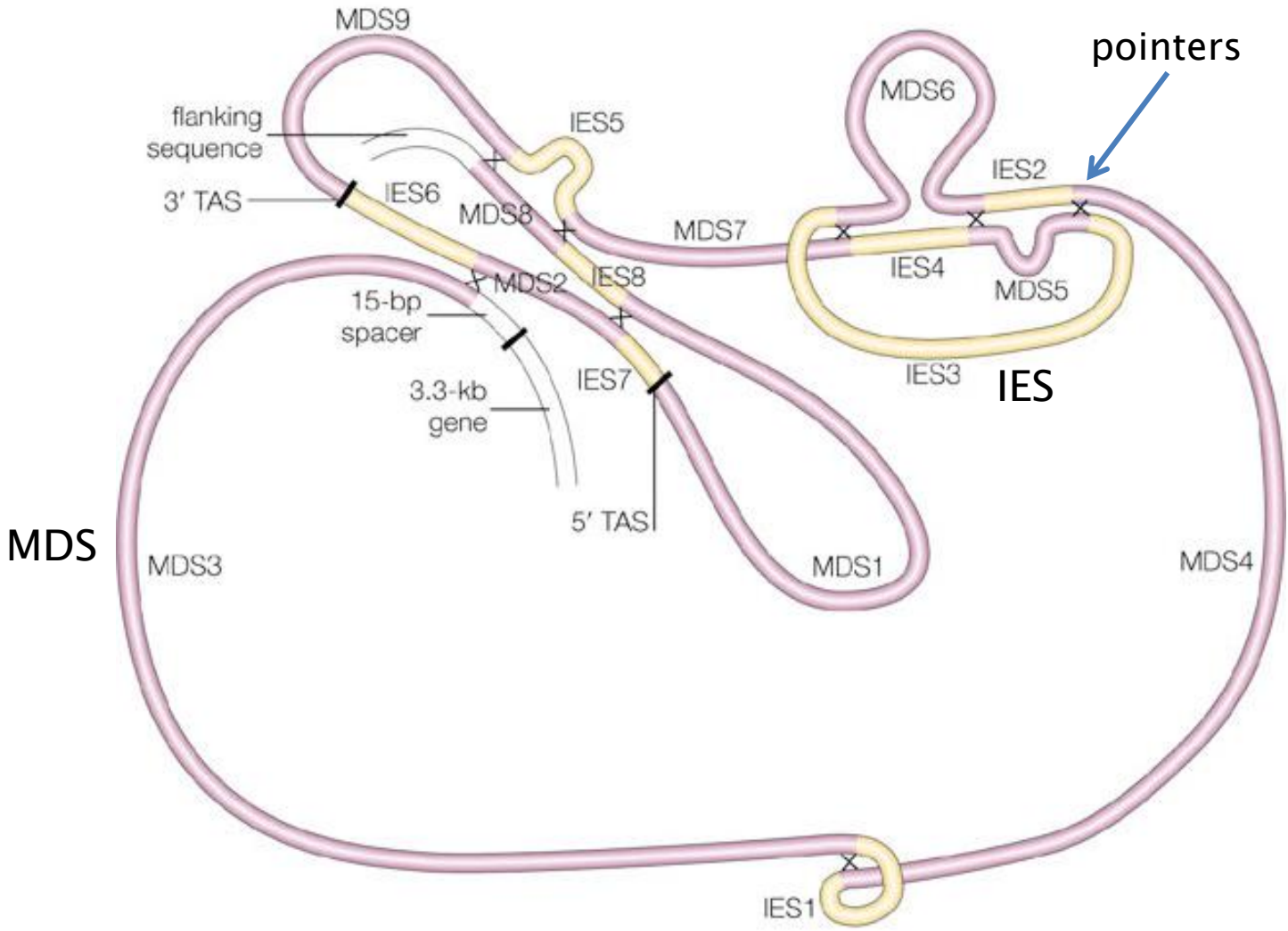
organizational features in their micronuclear and macronuclear genomes and carry out extensive developmental manipulations of their genomic DNA. These DNA phenomena, which considerably expand our view of genetic versatility, are the subject of this review.

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.

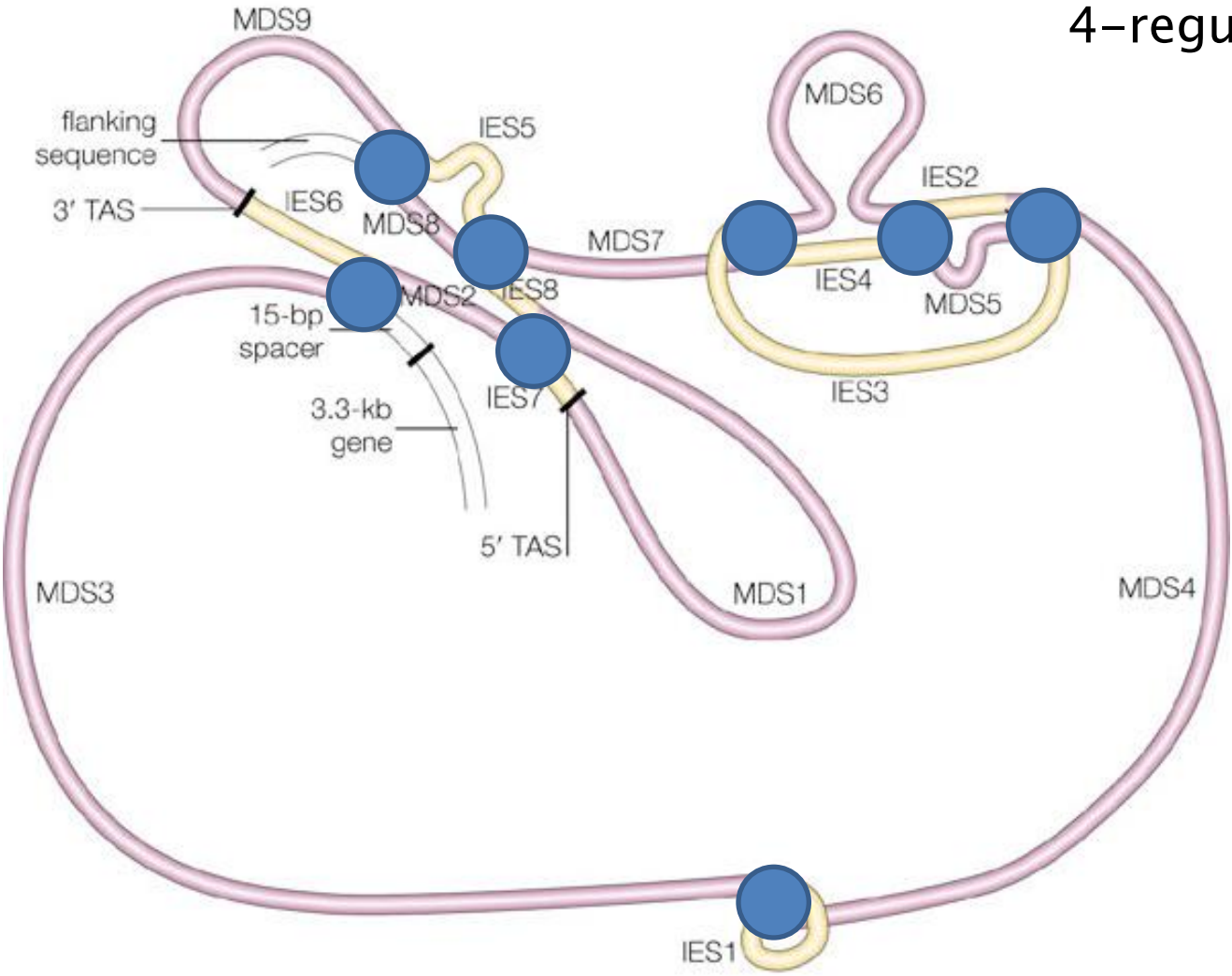
Micronuclear DNA consists of extremely long molecules typical of eukaryotic chromosomes. Genes are scattered along this DNA and are separated by large

genome gymnastics

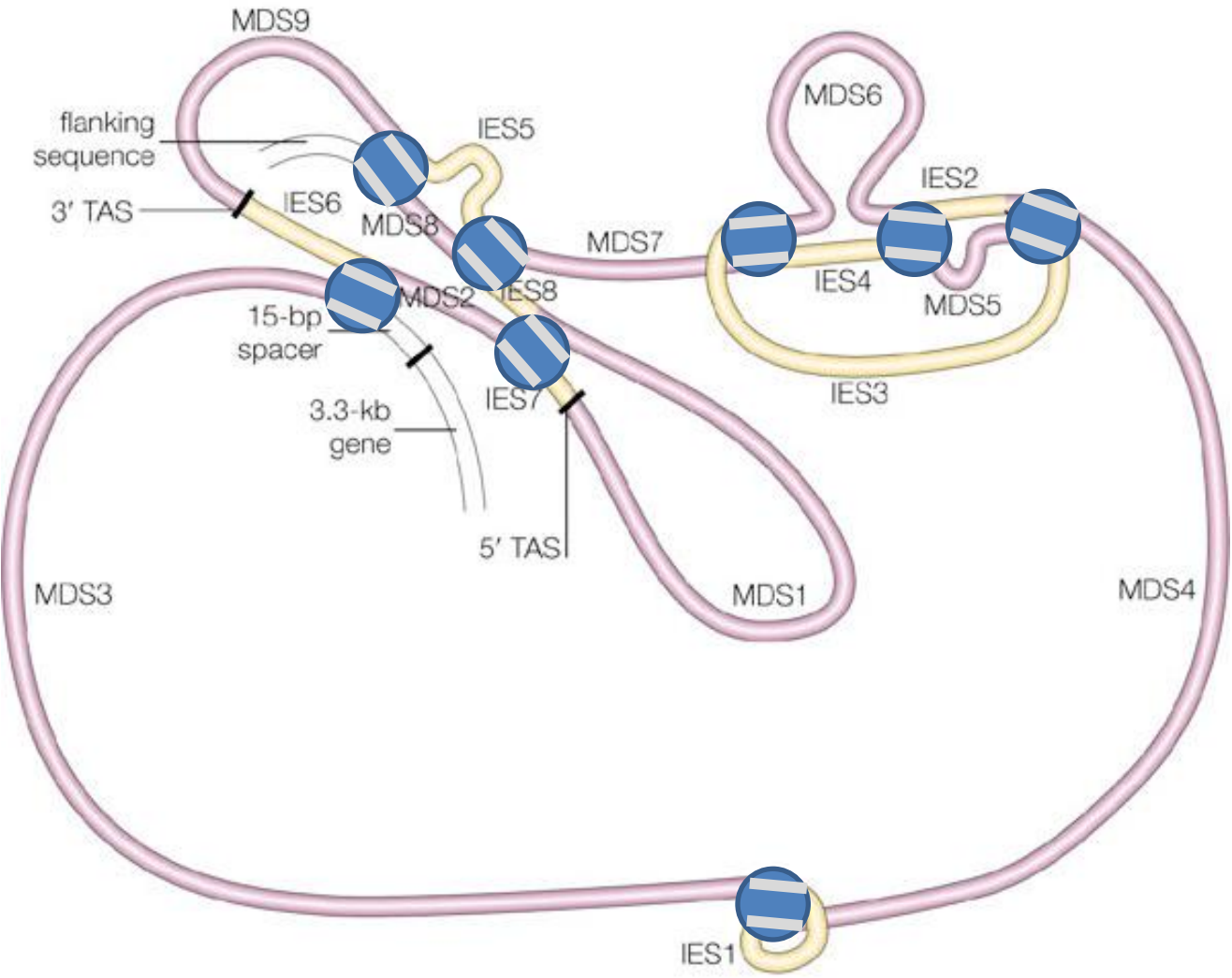


genome gymnastics

4-regular graph

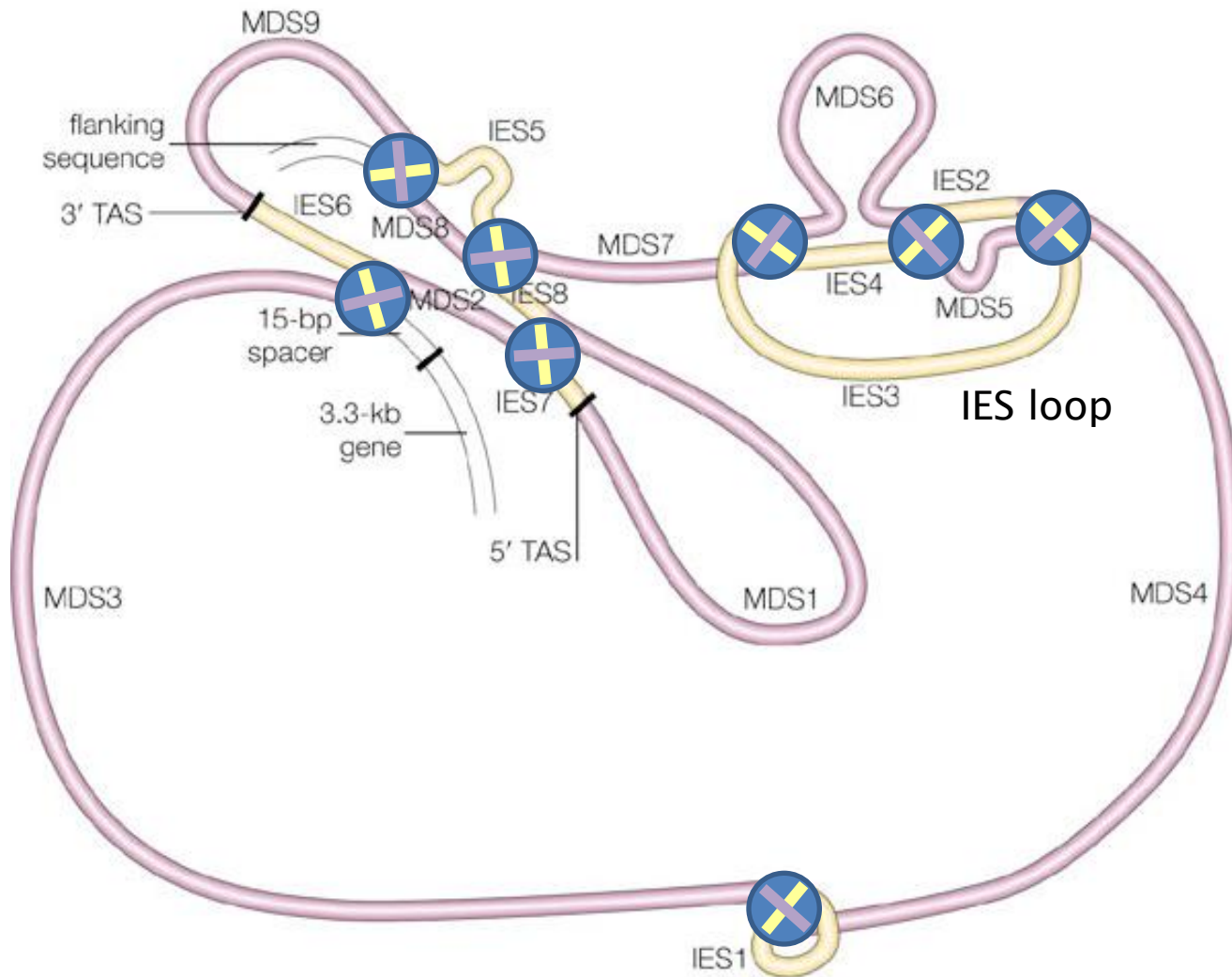


genome gymnastics



MIC

MDS and IES alternate



MAC

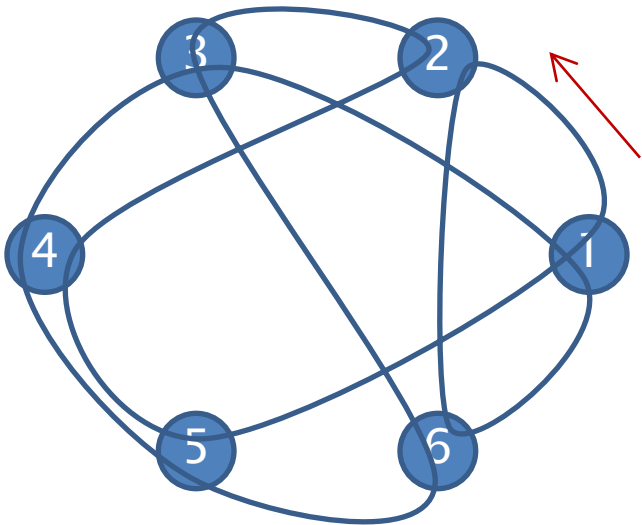
MDS and IES separated

How to study intramolecular gene assembly?

Eulerian circuits in 4-regular graphs

string = Euler circuit

126134563245



4-regular graph + Euler circuit

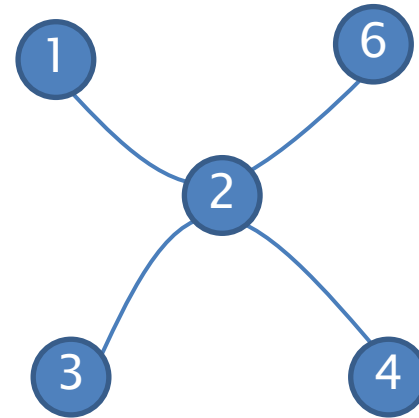
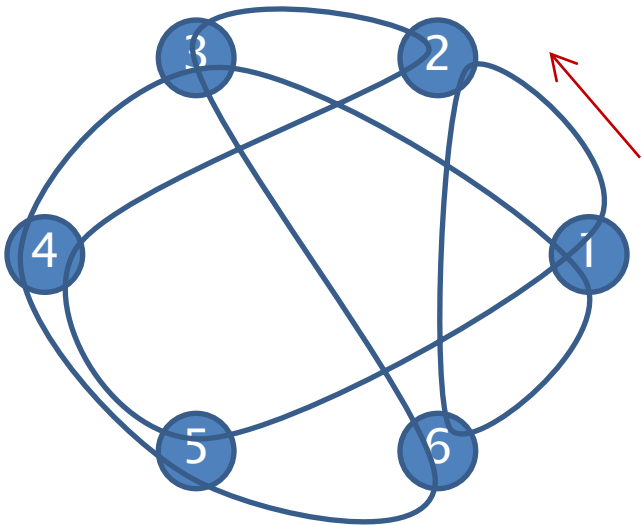
Eulerian circuits in 4-regular graphs

three transitions at 2

string = Euler circuit

126134563245

1-2-6 & 3-2-4



1-2-4 & 3-2-6

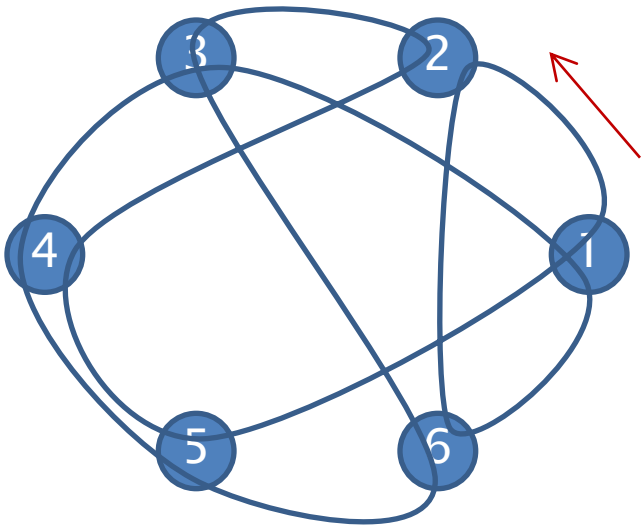
1-2-3 & 6-2-4

4-regular graph + Euler circuit

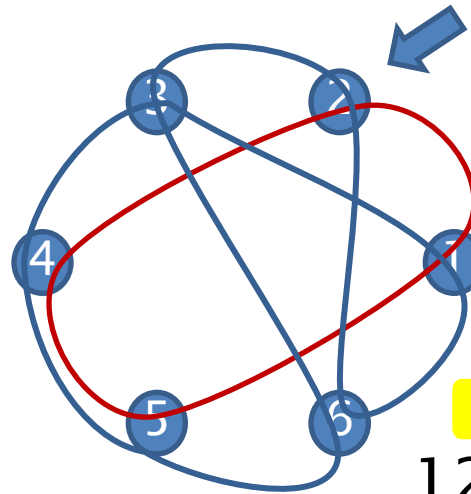
Eulerian circuits in 4-regular graphs

string = Euler circuit
 126134563245

1-2-6 & 3-2-4



4-regular graph + Euler circuit

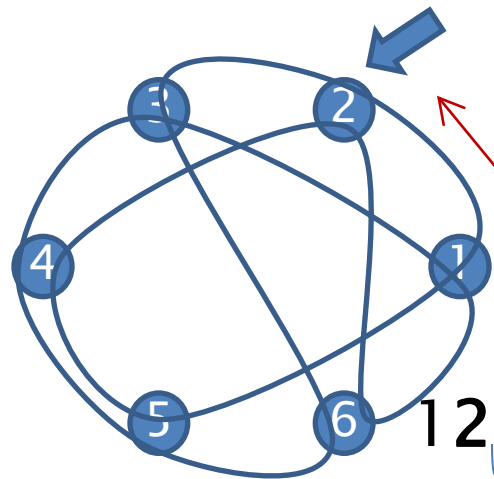


1-2-4 & 3-2-6

split

1245, 61345632

we can split **split** or **invert**



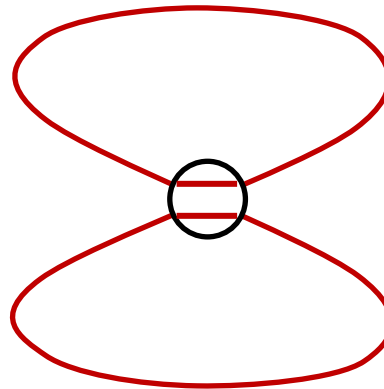
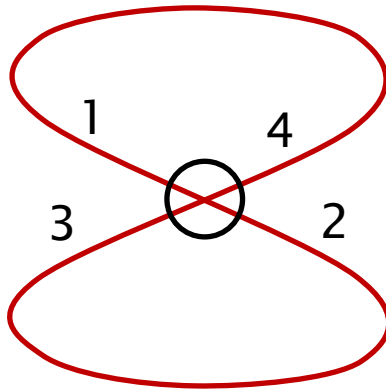
1-2-3 & 6-2-4

invert

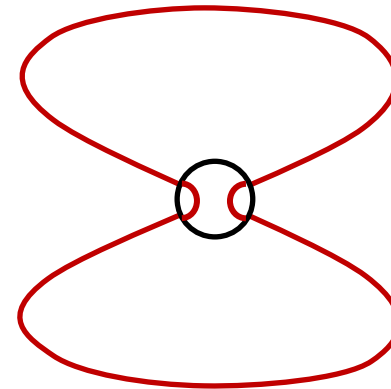
123654316245

reordering at vertices

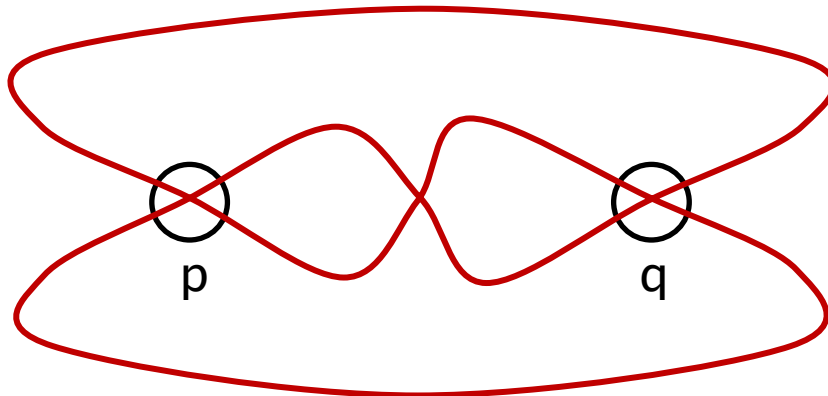
Kotzig 1966



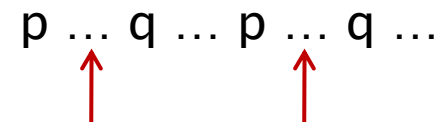
split



invert

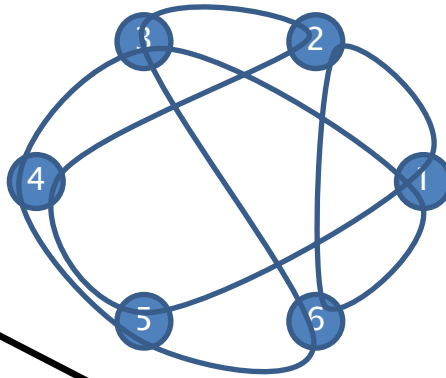
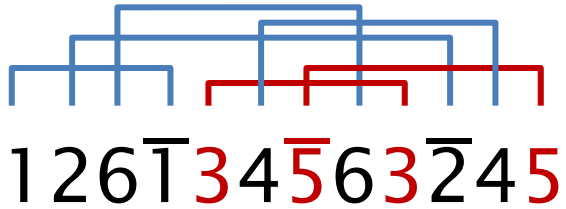


interlaced splits: swap

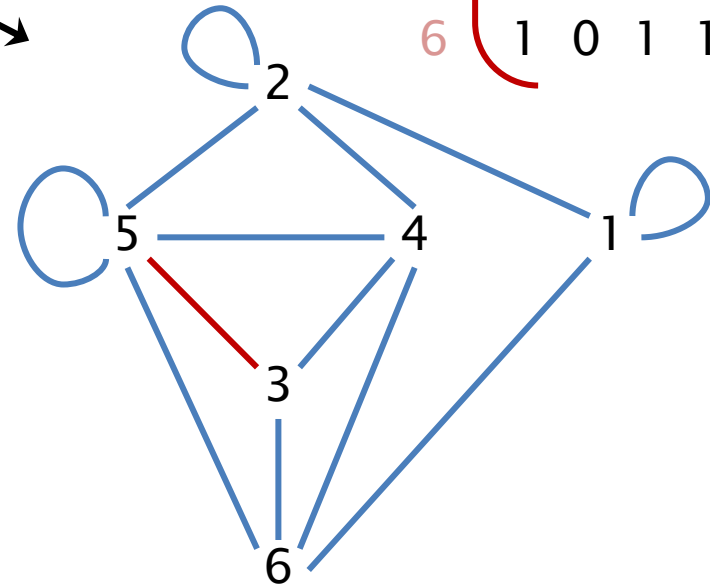
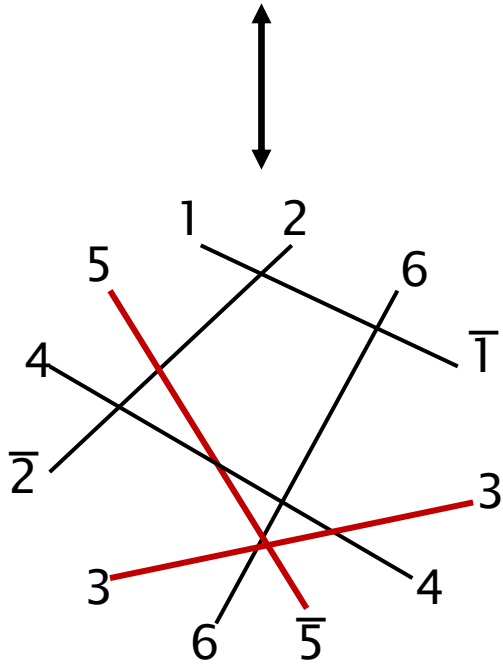


interlace / overlap graph

circle / interval graph



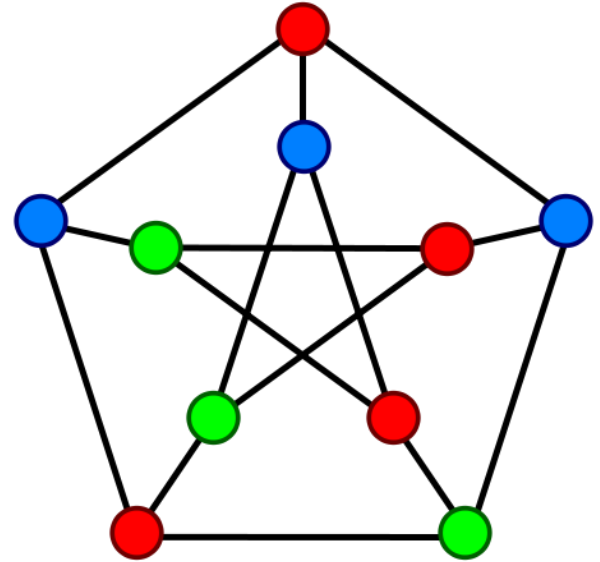
	1	2	3	4	5	6
1	1	1	0	0	0	1
2	1	1	0	1	1	0
3	0	0	0	1	1	1
4	0	1	1	0	1	1
5	0	1	1	1	1	1
6	1	0	1	1	1	0



interlace information

chromatic polynomial

“gene assembly polynomial”



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

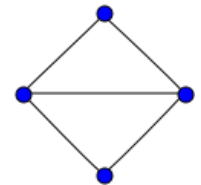
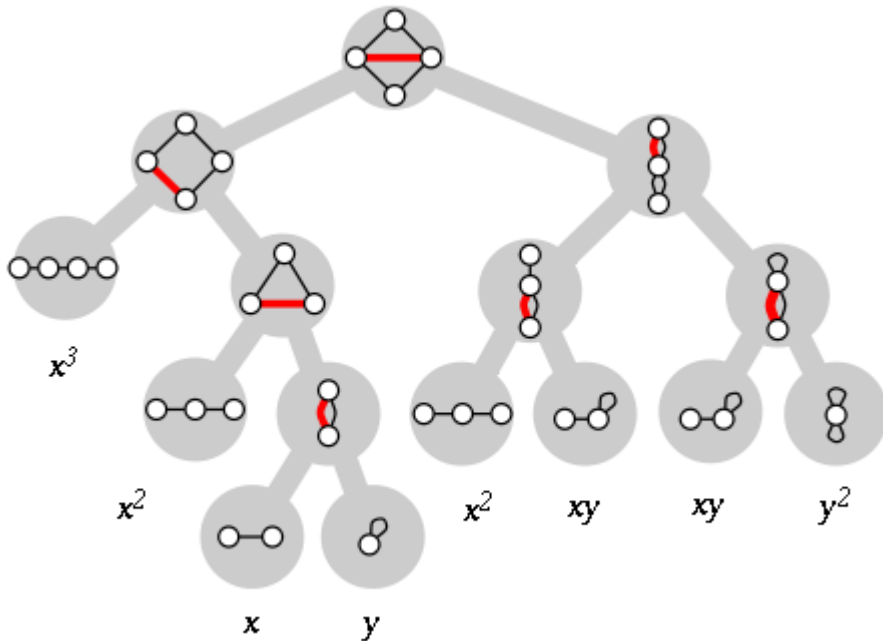
characteristic property

Tutte polynomial

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

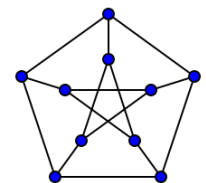
number components

of induced subgraph $G[A]$



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

deletion-contraction algorithm

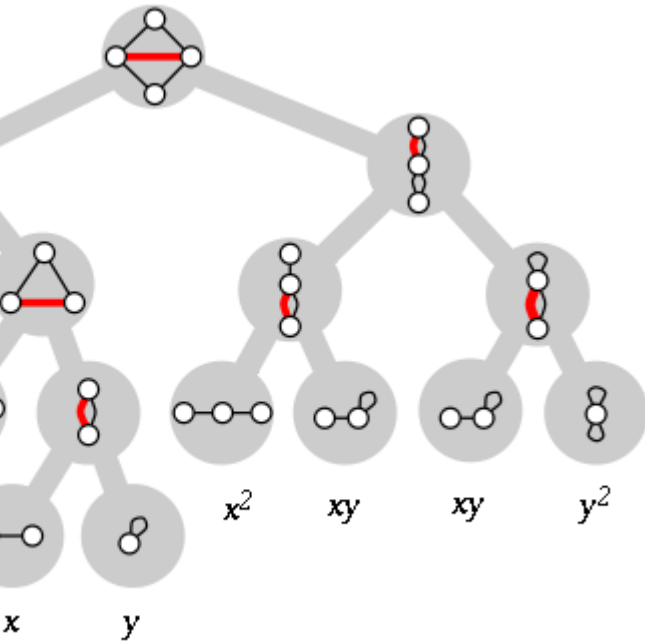


sage: P = graphs.PetersenGraph()

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

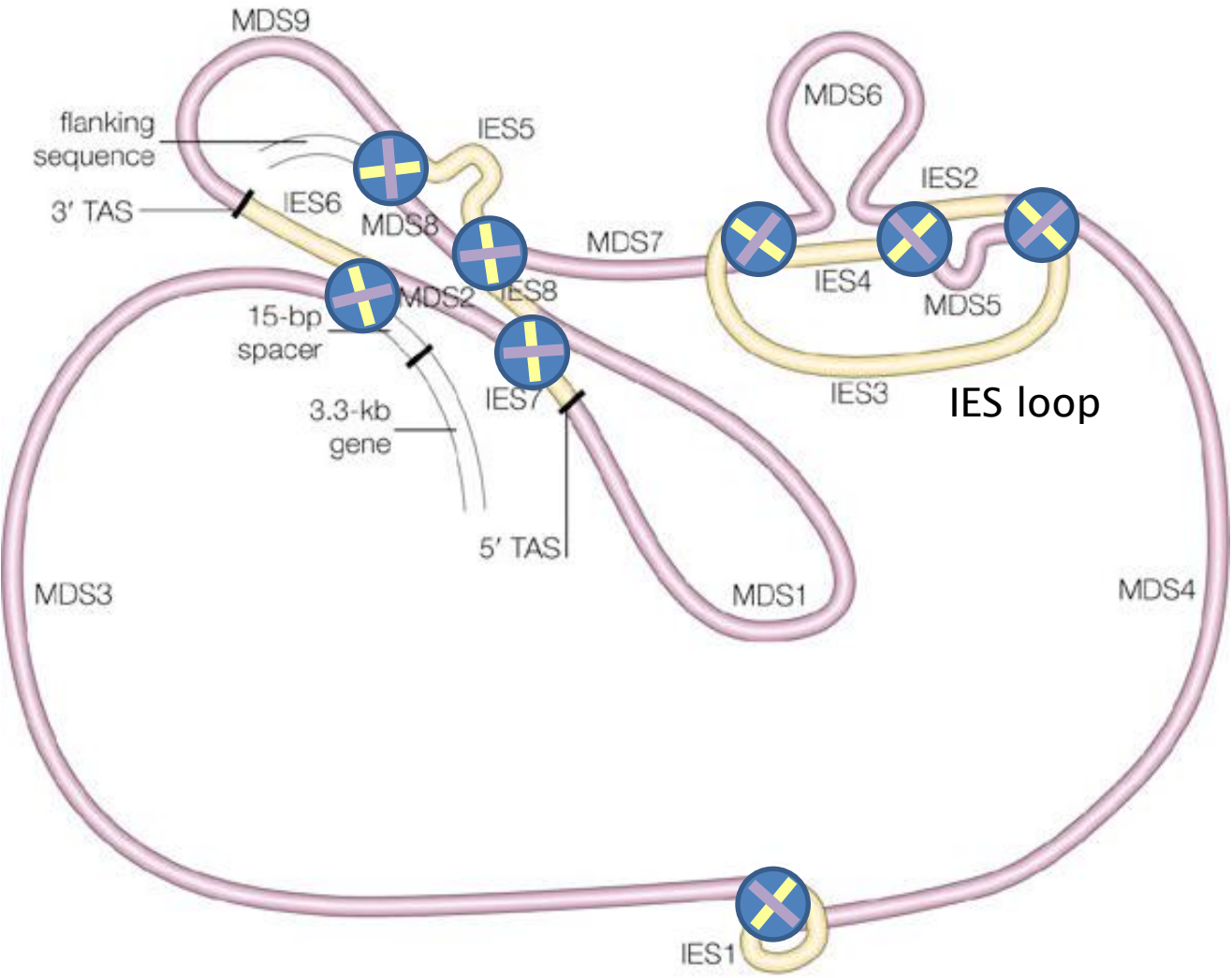
$T_G(1,1)$ # spanning forests

related

- knots “Jones polynomial”
- statistical physics

extended to matroids

genome gymnastics






MAC

MDS and IES separated

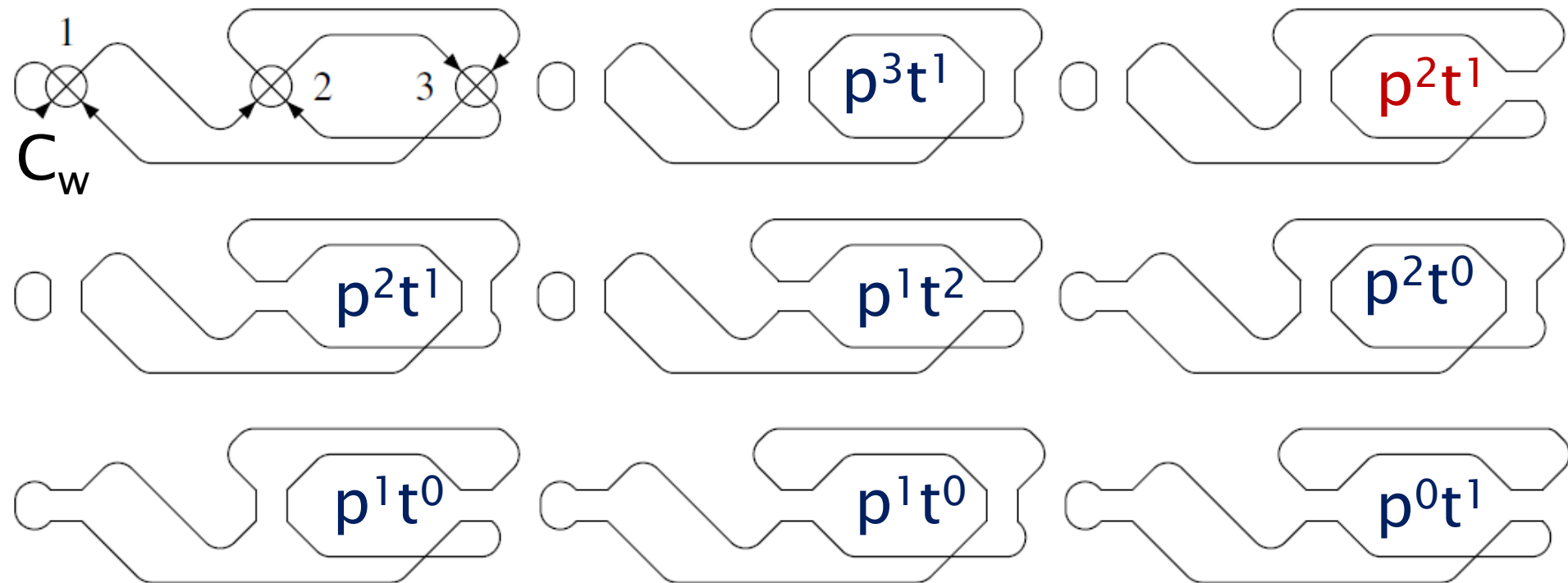
⇒ motivation: counting intermediate products

assembly polynomial

-  follows C
 -  consistent with C
 -  inconsistent with C
- } split
invert

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

2 consistent;
2 components



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

transition polynomials

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

$$\begin{aligned} \text{Diagram 1} &= \text{Diagram 2} - \text{Diagram 3} - \text{Diagram 4} - \text{Diagram 5} + \text{Diagram 6} + \text{Diagram 7} + \text{Diagram 8} - \text{Diagram 9} \\ &= 3^3 - 3^2 - 3^2 - 3^2 + 3 + 3 + 3 - 3 \end{aligned}$$

Roger Penrose, Applications of Negative Dimensional Tensors

Theorem (Traldi–Cohn–Lempel)

- C Euler cycle in 4-regular graph
- P circuit partition, transitions wrt C

D_1 follows C

split D_2 consistent with C

invert D_3 inconsistent with C

then

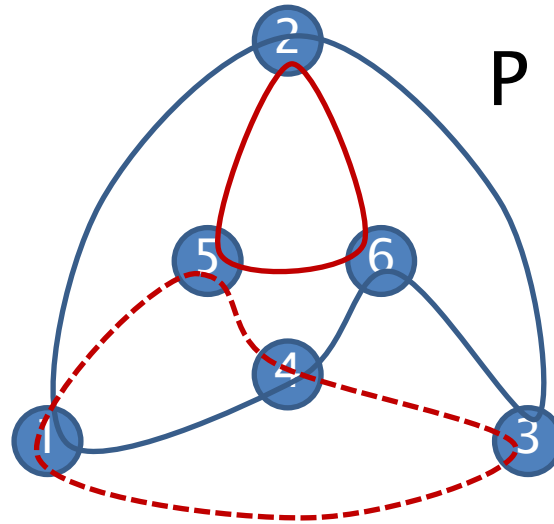
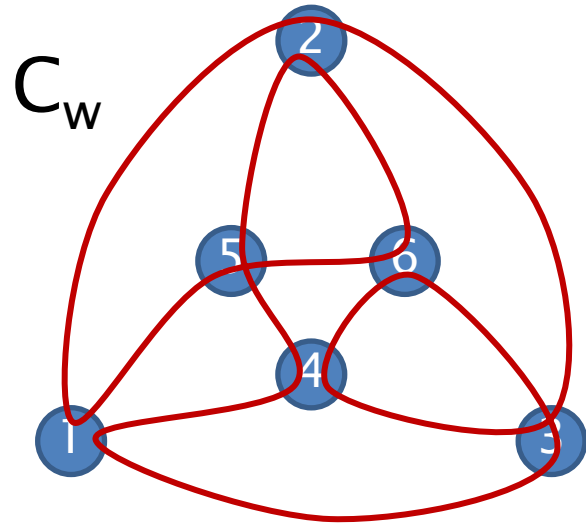
$$|P| = \nu \left(\underbrace{(I(C) + \Delta(D_3))}_{\text{overlap/interlace graph}} \setminus \underbrace{D_1}_{\text{deletion}} \right) + 1$$

nullity

loops /
diagonal matrix

4-regular graph from doc-string

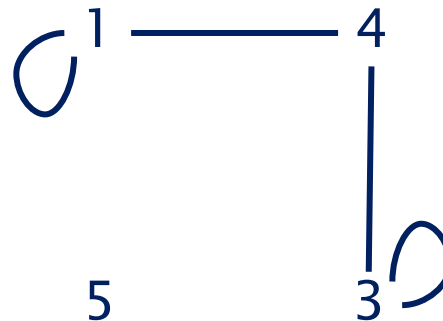
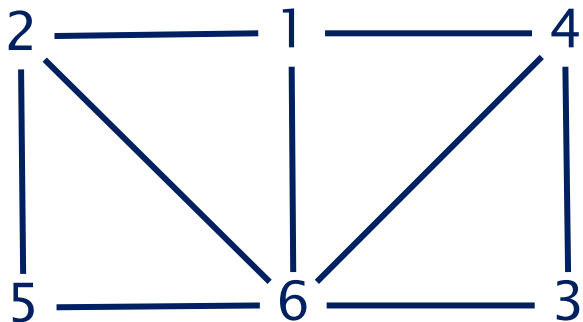
$w = 1\ 4\ 5\ 2\ 6\ 5\ 1\ 2\ 3\ 4\ 6\ 3$



- $D_1 = \{ 2, 6 \}$
- $D_2 = \{ 4, 5 \}$ split
- $D_3 = \{ 1, 3 \}$ invert

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

$I(C_w)$



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

rank=2 nullity = 2

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

combinatoriek
“tellen”

lineaire algebra
“matrices”

dankuwel ...



klaar