

O  
Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

# O Sekwencjach DNA i grafach

Bartek Wilczyński

27. lutego 2018

- Strona wykładu  
[regulomics.mimuw.edu.pl/wp/categories/wbo](http://regulomics.mimuw.edu.pl/wp/categories/wbo)
- e-mail [bartek@mimuw.edu.pl](mailto:bartek@mimuw.edu.pl)
- konultacje: środy 8:30–10:00, pokój 5770

- Projekty zaliczeniowe: 10+20 punktów
- Kolokwium 20 punktów
- Zaliczenie Ćwiczeń: > 30 punktów
- Egzamin ustny (wymagane zaliczenie ćwiczeń)

- Sekwencje biologiczne (DNA, RNA i białka)
- Podobieństwo sekwencji a ewolucja
- Porównywanie sekwencji biologicznych (miary, statystyki i algorytmy)
- Drzewa filogenetyczne - konstrukcja i zastosowania
- Sekwencje białkowe a ukryte modele Markowa
- Sekwencje nie- kodujące w genomach i motywy sekwencyjne
- Modele grafowe w biologii

O  
Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

- Computational Molecular Biology (P. Pevzner)
- Biological Sequence Analysis (R. Durbin i in.)
- Sequence – Evolution – Function (Koonin i Galperin)

# DNA and its role

O  
Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

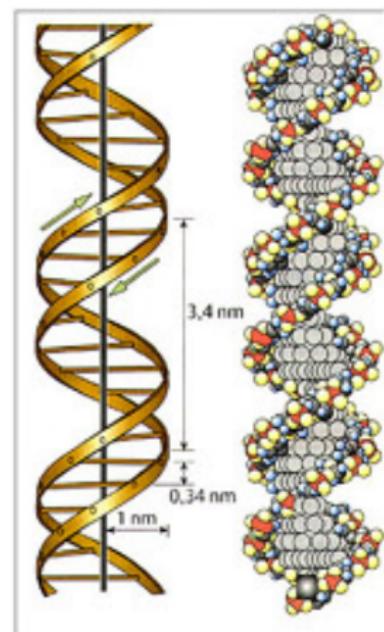
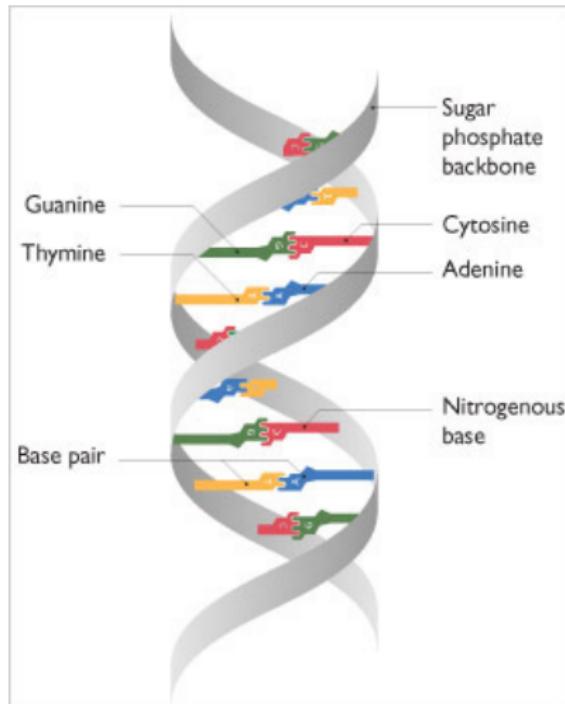


image (c) T. Maxwell

# DNA replication enables inheritance

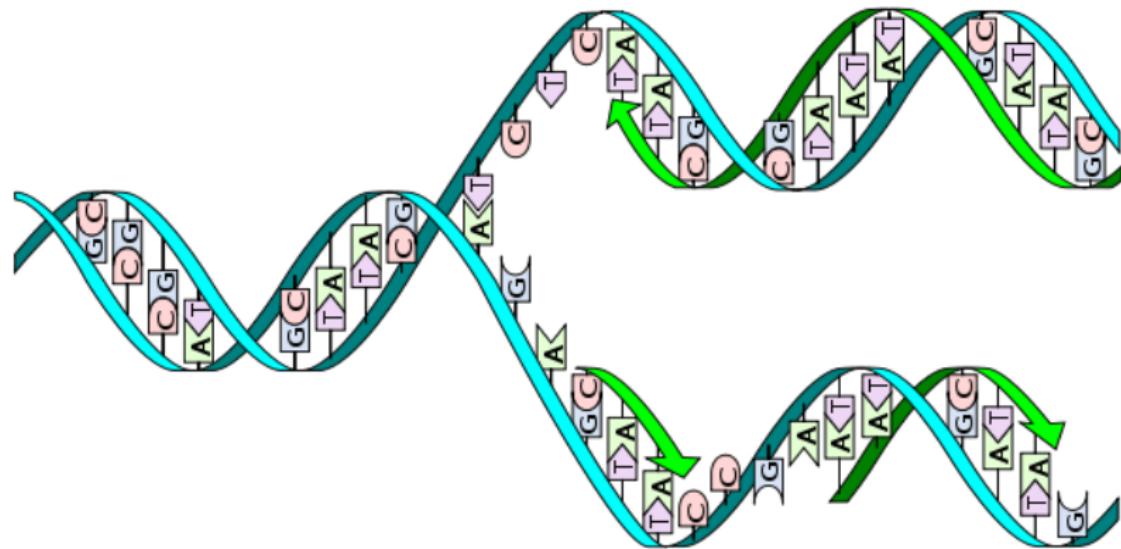


image (c) wikipedia

O

Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

# Central dogma of molecular biology

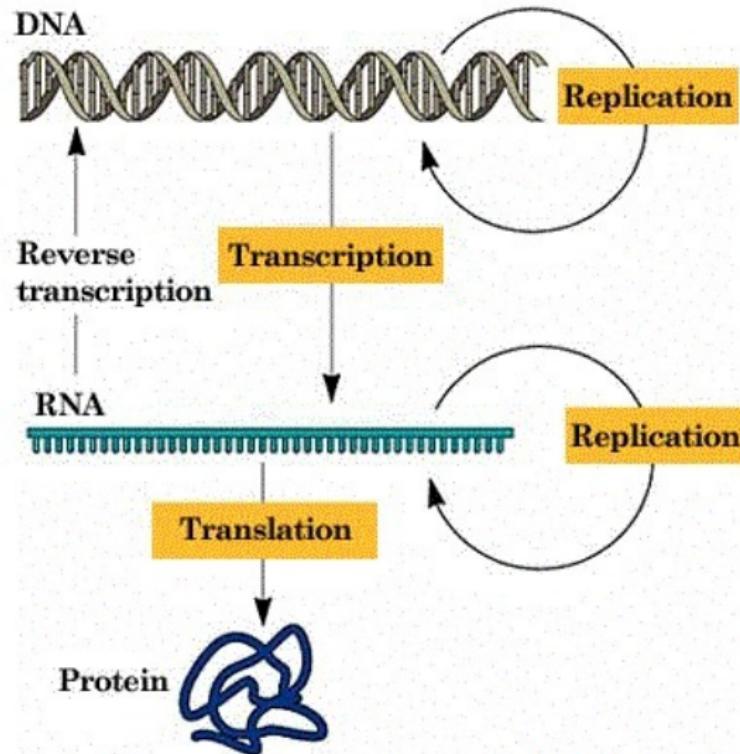


image (c) SCF IIT Delhi

# Microarray concept

O  
Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

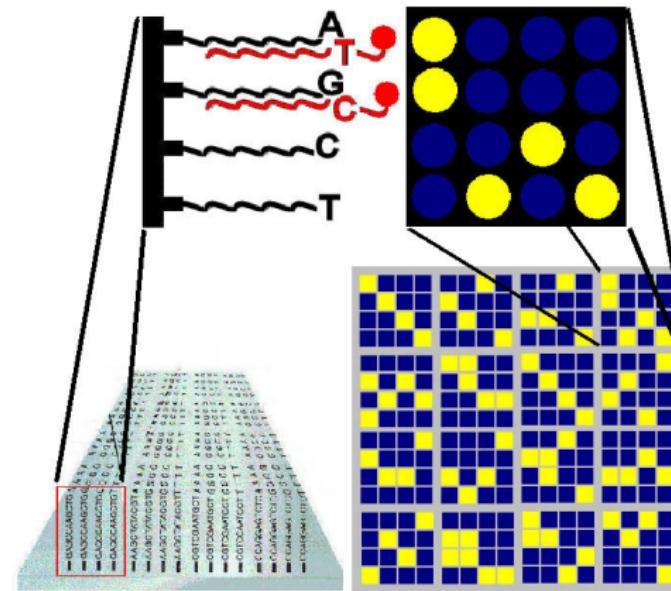


image (c) Steven M. Carr

O  
Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

affymetrix chip



image (c) Molecularstation.com

O

Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

# Sequencing by hybridization

DNA ARRAY C(4)															
AA	AT	AG	AC	TA	TT	TG	TC	GA	GT	GG	GC	CA	CT	CG	CC
AA															
AT		ATAG													
AG															
AC										ACGC					
TA									TAGG						
TT															
TG															
TC															
GA															
GT															
GG										GGCA					
GC	GCAA														
CA	CAAA														
CT															
CG															
CC															

DNA target TATCCGTTT (complement of ATAGGCAAA)

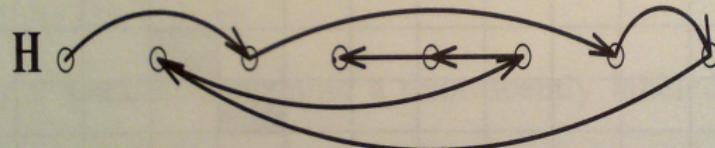
hybridizes to the array of all 4-mers:

A T A G G C A A A  
 A T A G  
 T A G G  
 A G G C  
 G G C A  
 G C A A  
 C A A A

image (c) P. Pevzner

## Sequence reconstruction (Hamiltonian path approach)

$$S = \{ \text{ATG, AGG, TGC, TCC, GTC, GGT, GCA, CAG} \}$$



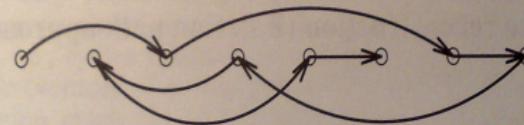
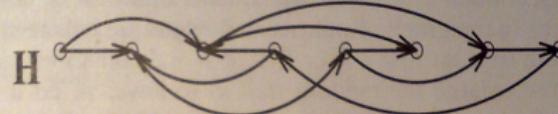
Vertices:  $l$ -tuples from the spectrum  $S$ . Edges: overlapping  $l$ -tuples.

Path visiting ALL VERTICES corresponds to sequence reconstruction      ATGCAGGTCC

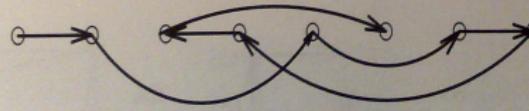
image (c) P.Pevzner

Multiple sequence reconstructions (Hamiltonian path approach)

$$S = \{ \text{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT} \}$$



ATGCGTGGCA



ATGGCGTGCA

# Hamiltonian and Eulerian graphs

Finding Hamiltonian paths is NP-complete, while finding Eulerian paths is easy.

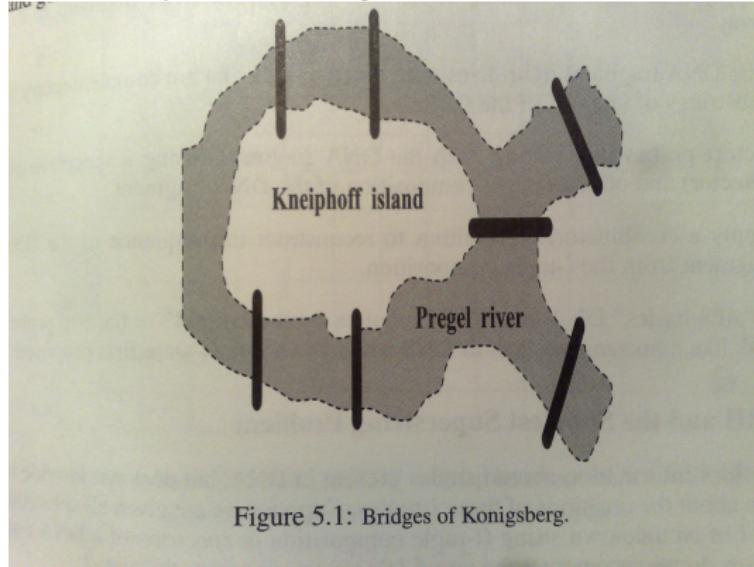


Figure 5.1: Bridges of Konigsberg.

image (c) P.Pevzner

## SBH - Eulerian formulation

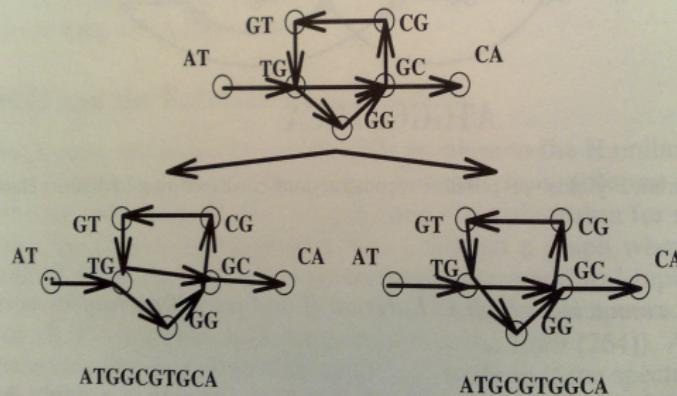
 $S = \{ATG, TGG, TGC, GTG, GGC, GCA, GCG, CGT\}$ Vertices correspond to  $(l-1)$ -tuples.Edges correspond to  $l$ -tuples from the spectrum

image (c) P.Pevzner

# Other uses of microarrays

O  
Sekwencjach  
DNA i grafach

Bartek  
Wilczyński

Sprawy  
organizacyjne

Biological  
sequences

Microarrays

SBH

Microarray  
probe design

- SBH not practical due to hybridization errors, superseded by Next generation sequencing
- Gene arrays for RNA abundance quantification
- Snip arrays for detecting mutations (disease screening, paternity tests)
- aCGH arrays for detecting copy number variation
- biochips for quick pathogen detection

- We are given a set of  $n$  different DNA sequences (targets)  
 $\mathcal{S} = \{s_1 \dots s_n\}$
- We need to design a set  $\mathcal{P} = \{p_1 \dots p_n\}$  of  $n$  sequences (probes) of length  $k$ , such that for each  $i \in \{1 \dots n\}$ , probe  $p_i$  hybridizes with sequence  $s_i$  and does not hybridize with any other sequence  $s_j, j \neq i$ .
- Depending on the amount of sequence identity and parameter  $k$ , there might be no valid solutions or exponentially many solutions
- Instead of searching for probes of the same length, one might search for probes of the same *melting temperature*

$$t_m = 4 \cdot \#GC + 2 \cdot \#AT$$