

Ewolucja współczesnej biomedycyny w kierunku informatyki

Tomasz Gambin



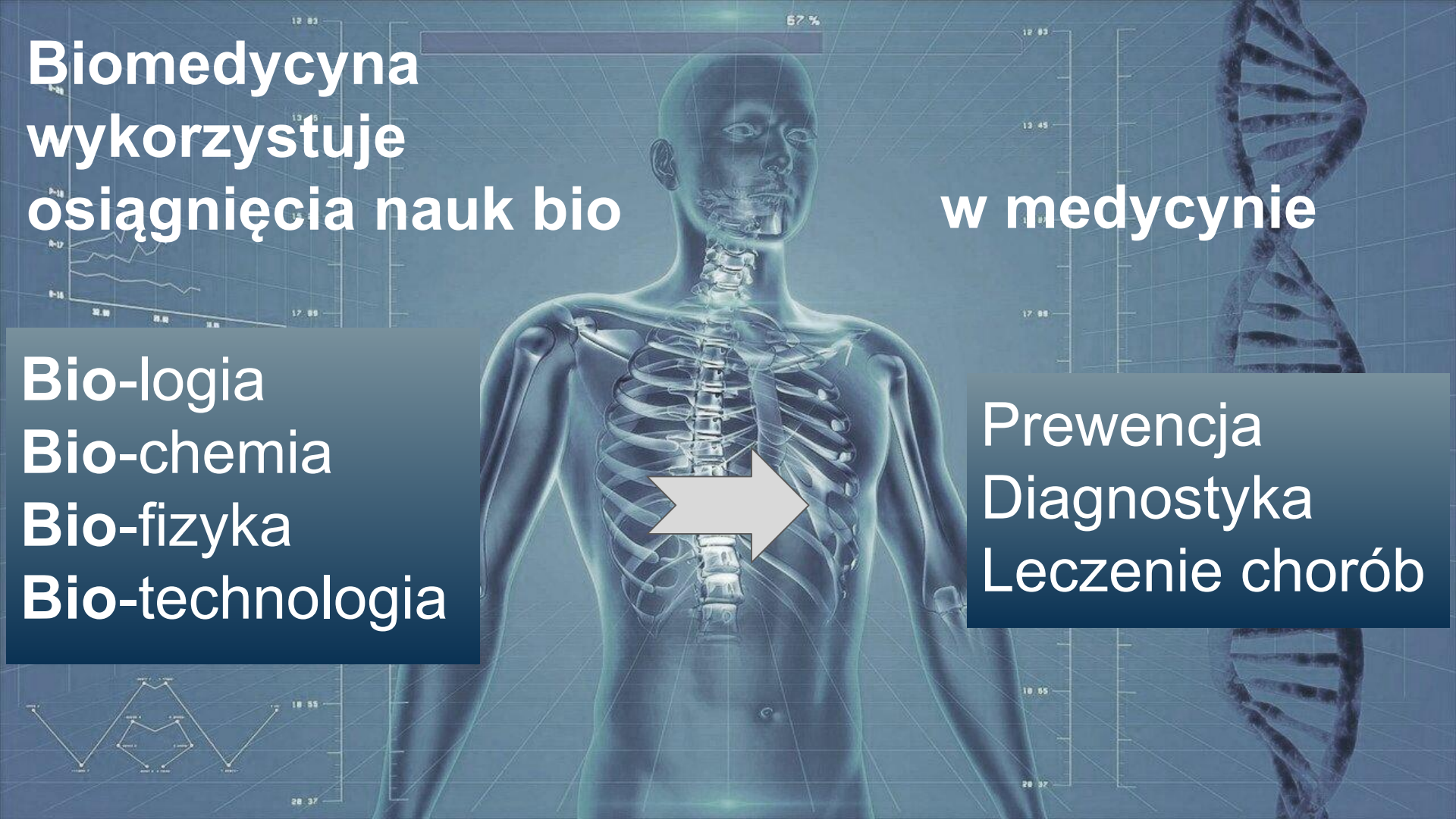
**Wydział Elektroniki
i Technik Informatycznych**

**Biomedycyna
wykorzystuje
osiągnięcia nauk bio**

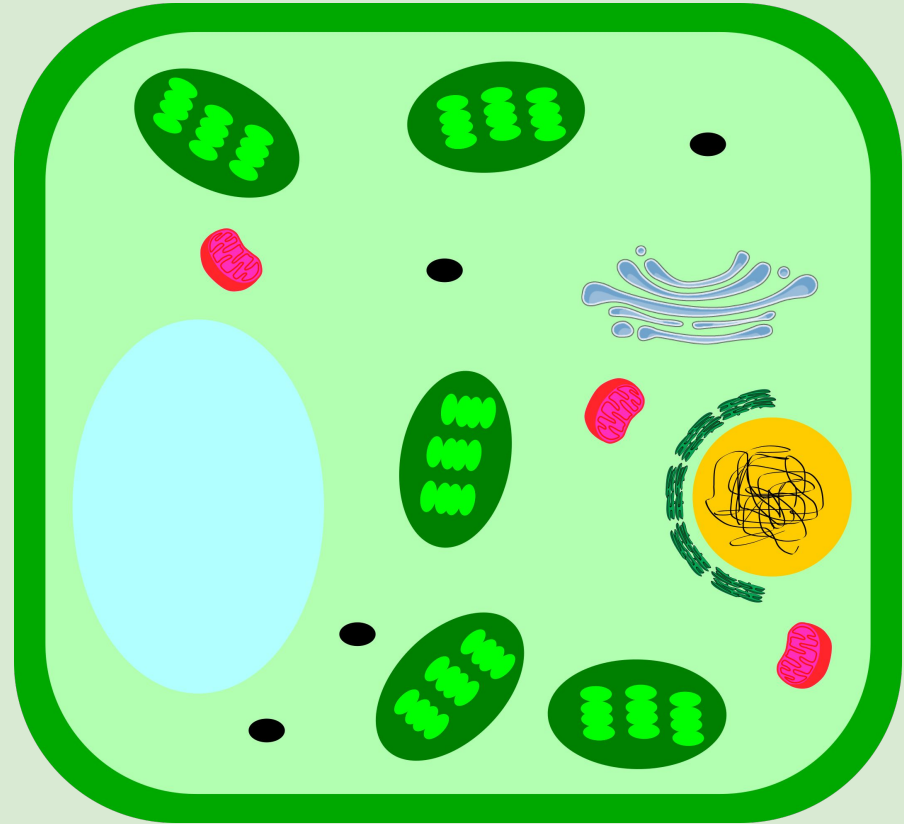
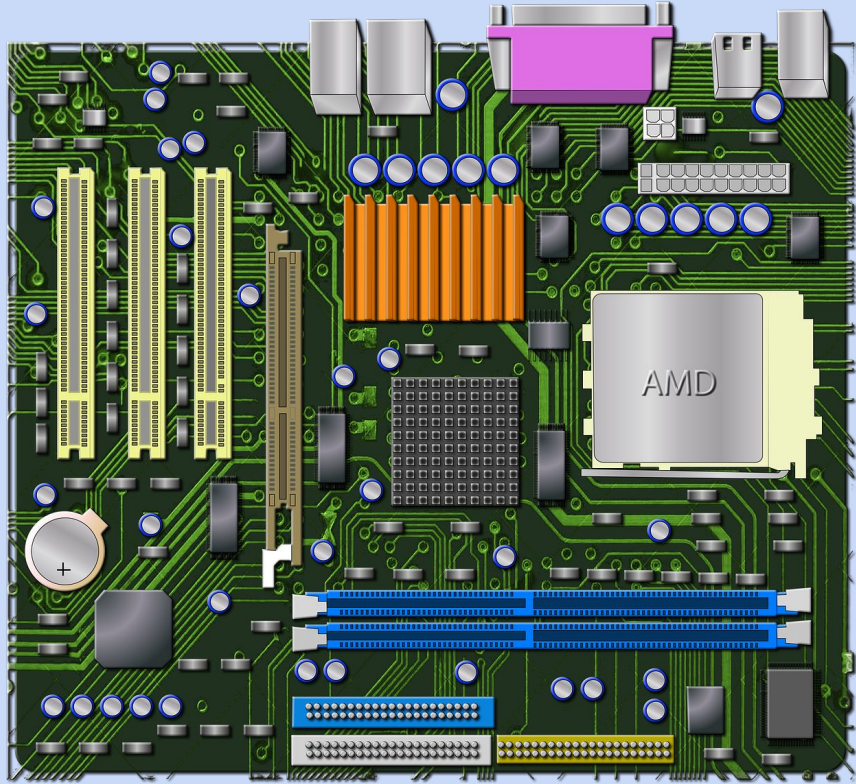
w medycynie

**Bio-logia
Bio-chemia
Bio-fizyka
Bio-technologie**

**Prewencja
Diagnostyka
Leczenie chorób**



Co mają wspólnego informatyka i nauki biologiczne?



Cyfrowy zapis informacji w komputerze i w żywym organizmie

1001000 1000101 1001100
1001100 1001111 0100000
1010111 1001111 1010010
1001100 1000100

alfabet binarny

ATGGTGGCAGGGAAGGCTGAGCCA
GCCATGCCA

alfabet czteroliterowy

Kodowanie informacji - kod ASCII

1001000 1000101 1001100
1001100 1001111 0100000
1010111 1001111 1010010
1001100 1000100

1000001	A	1001110	N
1000010	B	1001111	O
1000011	C	1010000	P
1000100	D	1010001	Q
1000101	E	1010010	R
1000110	F	1010011	S
....

Kodowanie informacji - kod ASCII

1001000 → H

1000101 → E

1001100 → L

1001100 → L

1001111 → O

0100000 →

1010111 → W

1001111 → O

1010010 → R

1001100 → L

1000100 → D



HELLO WORLD

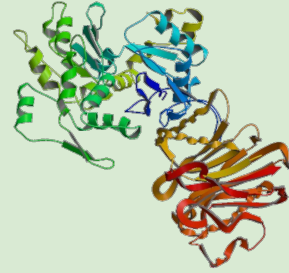
Kodowanie informacji - kod genetyczny

**ATGGTGGCAGGGGAAGGCTGAGCCA
GCCATGCCA**

	T	C	A	G	
T	TTT Phe	TCT Ser	TAT Tyr	TGT Cys	T
	TTC Phe	TCC Ser	TAC Tyr	TGC Cys	C
	TTA Leu	TCA Ser	TAA stop	TGA stop	A
	TTG Leu	TCG Ser	TAG stop	TGG Trp	G
C	CTT Leu	CCT Pro	CAT His	CGT Arg	T
	CTC Leu	CCC Pro	CAC His	CGC Arg	C
	CTA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CTG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	ATT Ile	ACT Thr	AAT Asn	AGT Ser	T
	ATC Ile	ACC Thr	AAC Asn	AGC Ser	C
	ATA Ile	ACA Thr	AAA Lys	AGA Arg	A
	ATG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GTT Val	GCT Ala	GAT Asp	GGT Gly	T
	GTC Val	GCC Ala	GAC Asp	GGC Gly	C
	GTA Val	GCA Ala	GAA Glu	GGA Gly	A
	GTG Val	GCG Ala	GAG Glu	GGG Gly	G

Kodowanie informacji - kod genetyczny

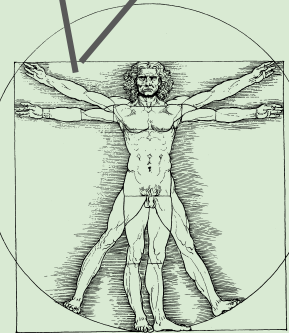
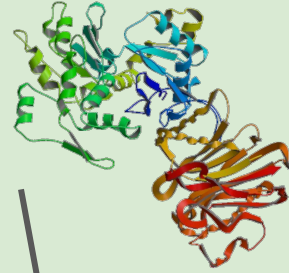
ATG	—————→	Met
GTG	—————→	Val
GCA	—————→	Ala
GGG	—————→	Gly
AAG	—————→	Lys
GCT	—————→	Ala
GAG	—————→	Glu
CCA	—————→	Pro
GCC	—————→	Ala
ATG	—————→	Met
CCA	—————→	Pro



Kodowanie informacji - kod genetyczny

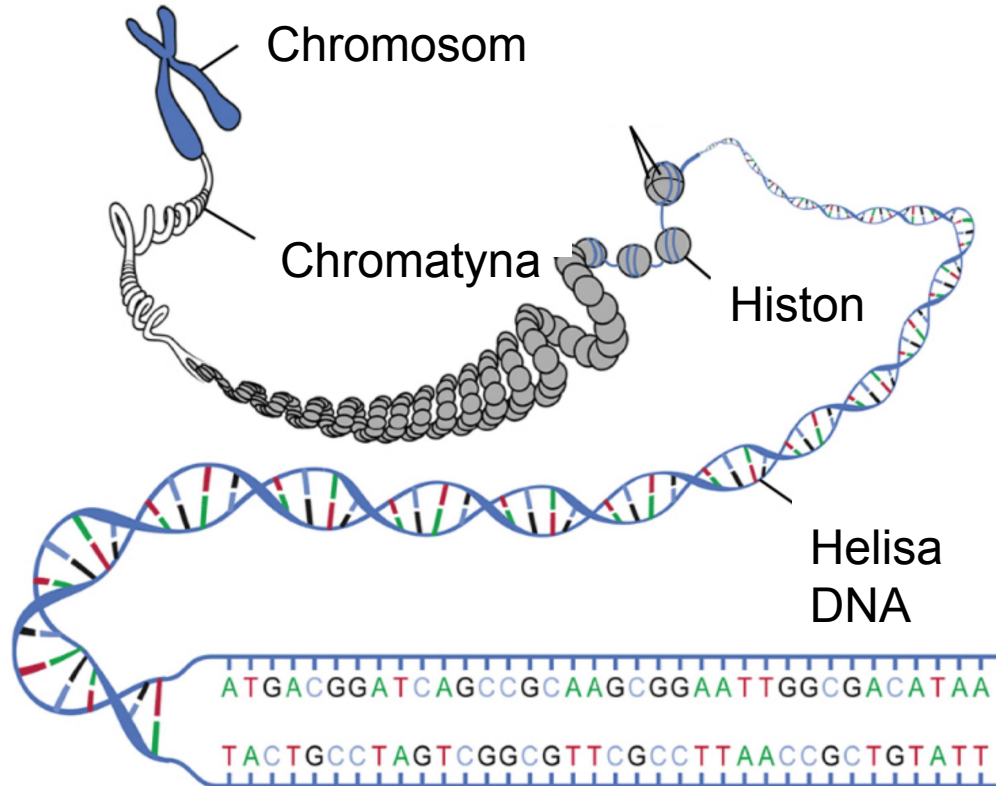
1001000	→	H
1000101	→	E
1001100	→	L
1001100	→	L
1001111	→	O
0100000	→	
1010111	→	W
1001111	→	O
1010010	→	R
1001100	→	L
1000100	→	D

ATG	→	Met
GTG	→	Val
GCA	→	Ala
GGG	→	Gly
AAG	→	Lys
GCT	→	Ala
GAG	→	Glu
CCA	→	Pro
GCC	→	Ala
ATG	→	Met
CCA	→	Pro



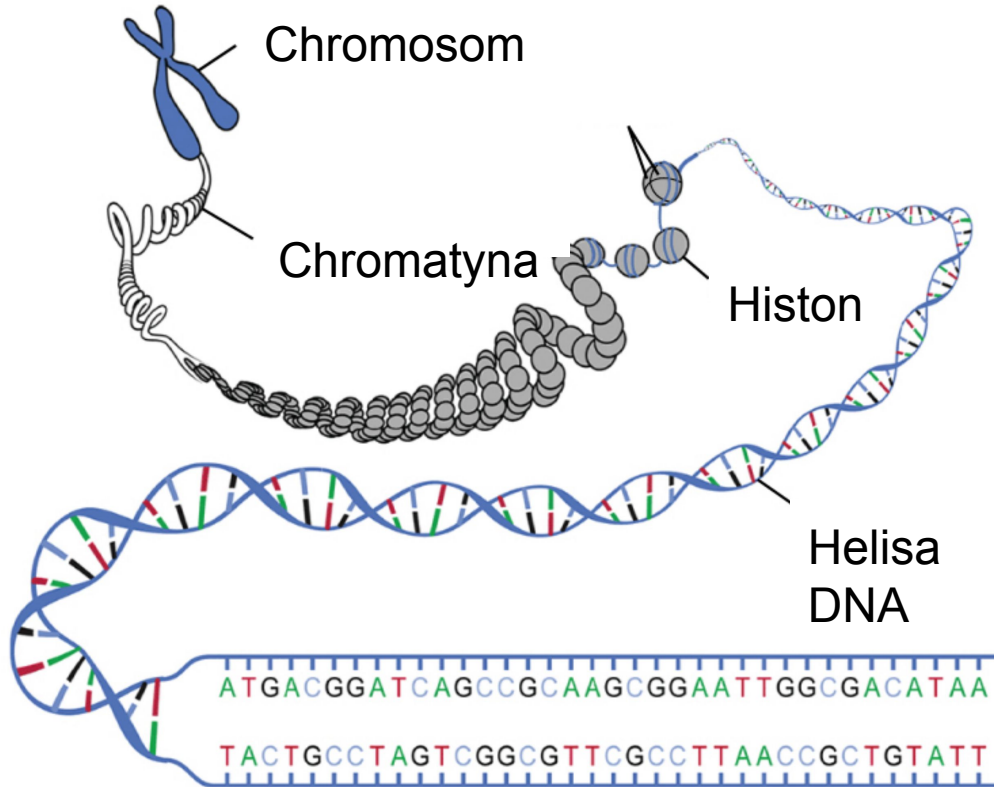
HELLO WORLD

Informacja w genomie



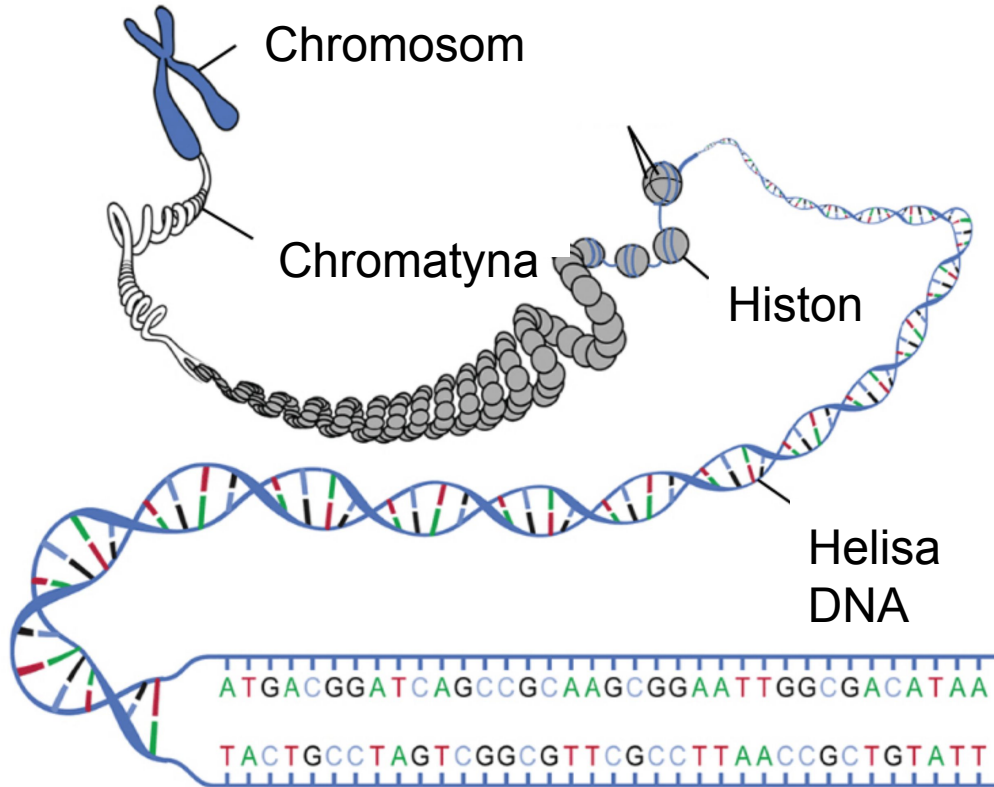
- 2x3 miliardy nukleotydów

Informacja w genomie



- 2x3 miliardy nukleotydów
- A: 00, C: 01, T: 10, G: 11
- 12 miliardów bitów, czyli 1500 MB

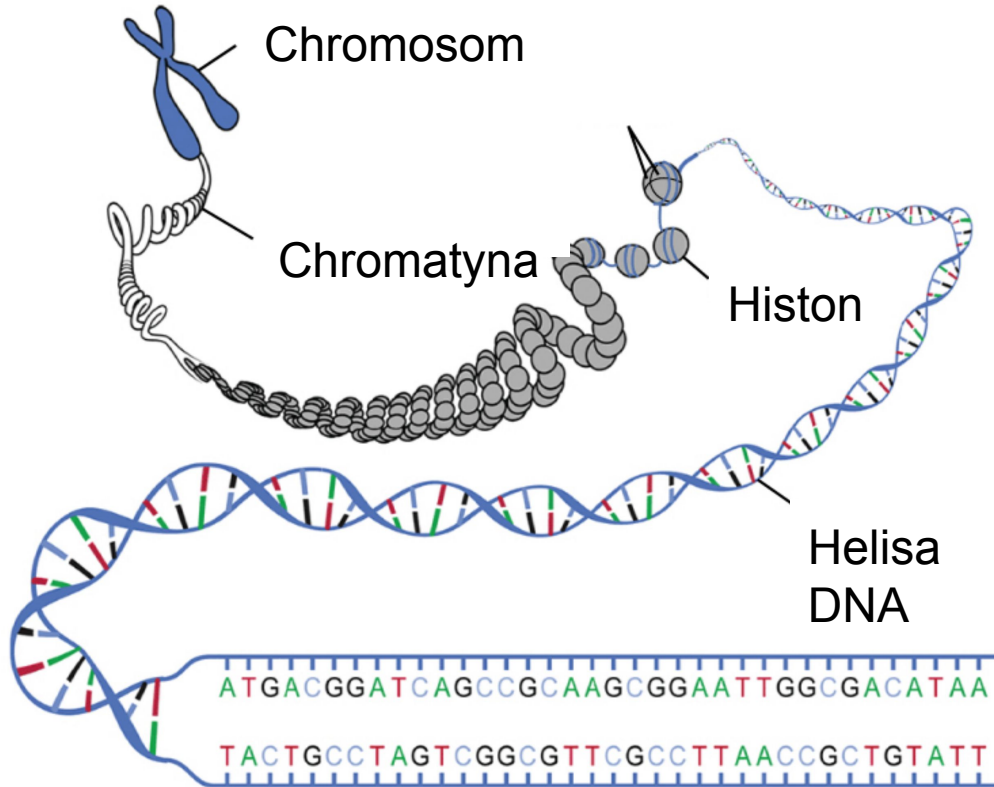
Informacja w genomie



- 2x3 miliardy nukleotydów
- A: 00, C: 01, T: 10, G: 11
- 12 miliardów bitów, czyli 1500 MB
- 1 cd rom ~ 750 MB



Informacja w genomie

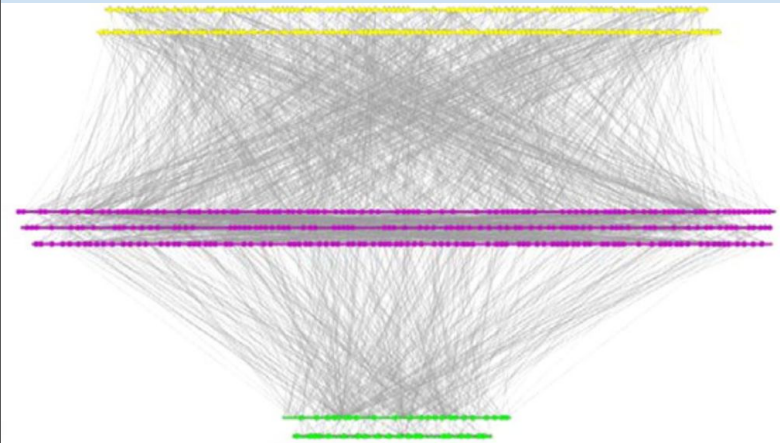


- 2x3 miliardy nukleotydów
- A: 00, C: 01, T: 10, G: 11
- 12 miliardów bitów, czyli 1500 MB
- 1 cd rom ~ 750 MB
- **Ale to nie wszystko:** człowiek ma 37 tera (10^{12}) komórek, w których DNA może się zmieniać w trakcie życia



Architektura modularna

Graf wywołań systemowych
w Linuksie

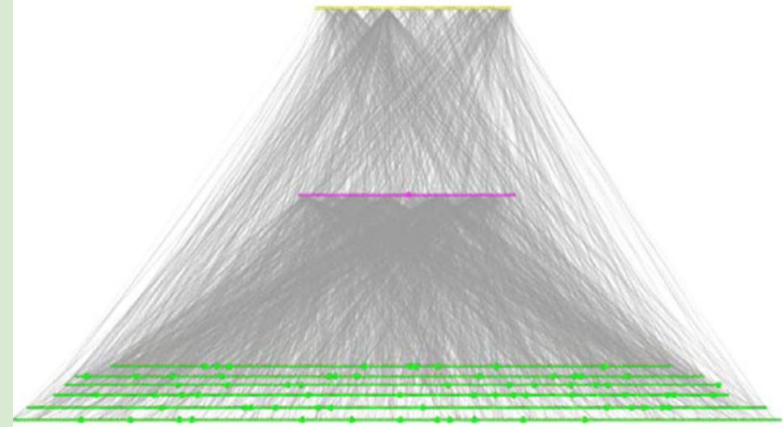


funkcje/geny
sterujące

zarządcy
średniego
szczebla

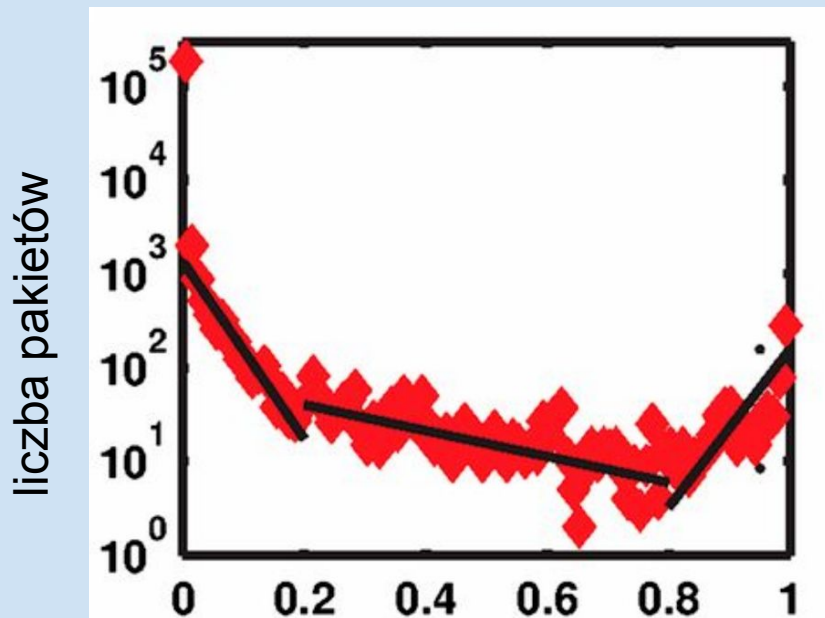
funkcje/geny
wykonawcze

Regulacja ekspresji genów
u bakterii *E. Coli*

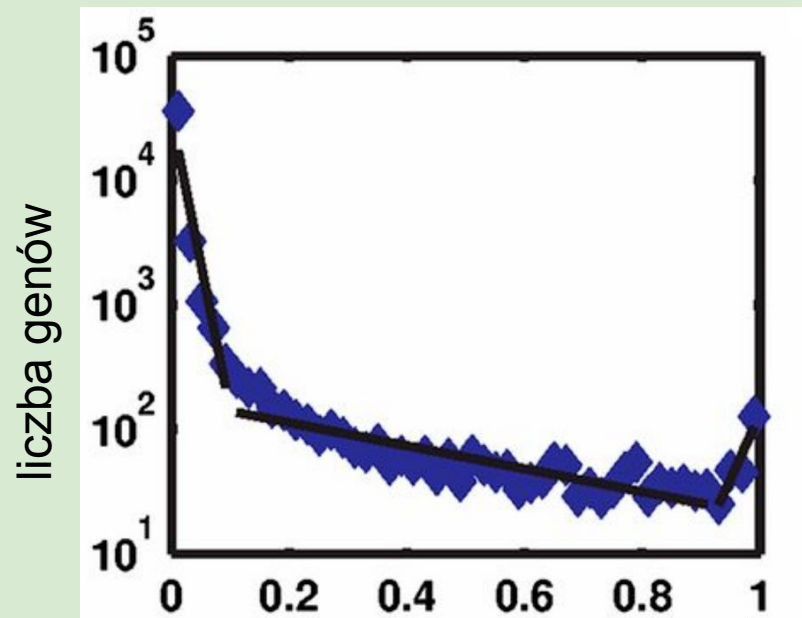


Źródło: Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. Yan *i in.* PNAS, 2010

Wykorzystanie modułów

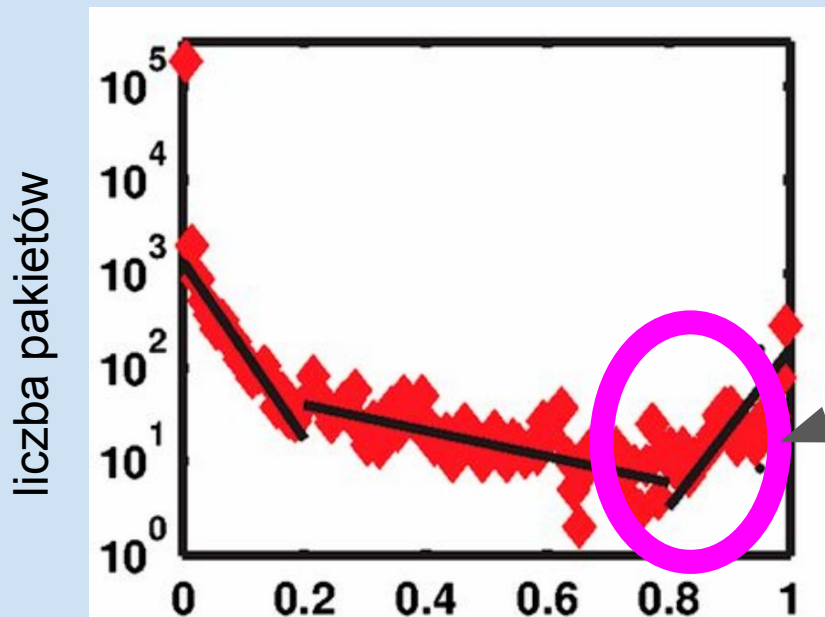


częstość instalacji danego pakietu na komputerach z systemem Linuks

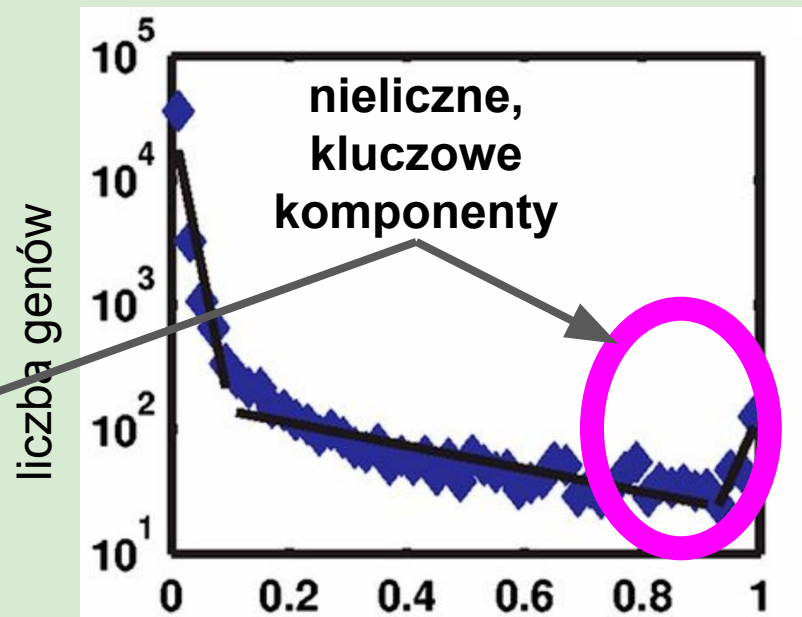


częstość występowania danego genu w genomach różnych gatunków bakterii

Wykorzystanie modułów

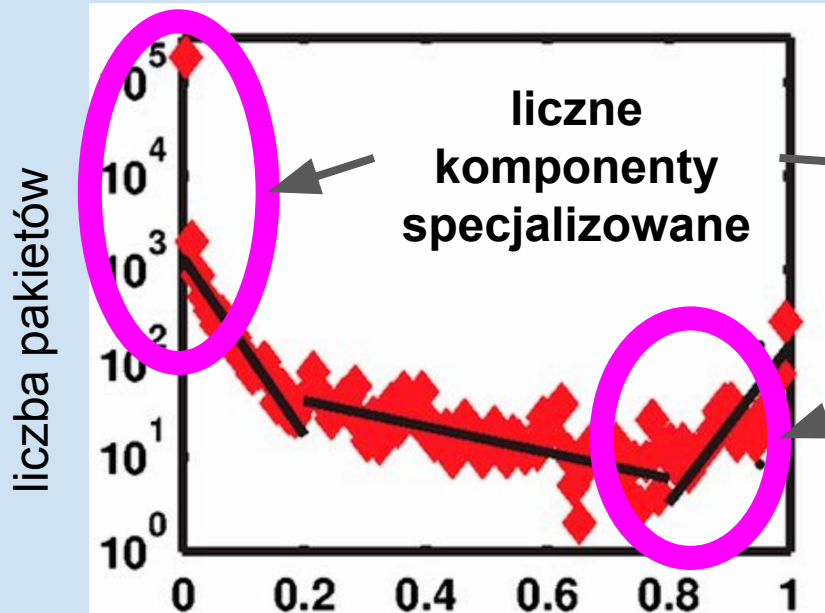


częstość instalacji danego pakietu na komputerach z systemem Linuks

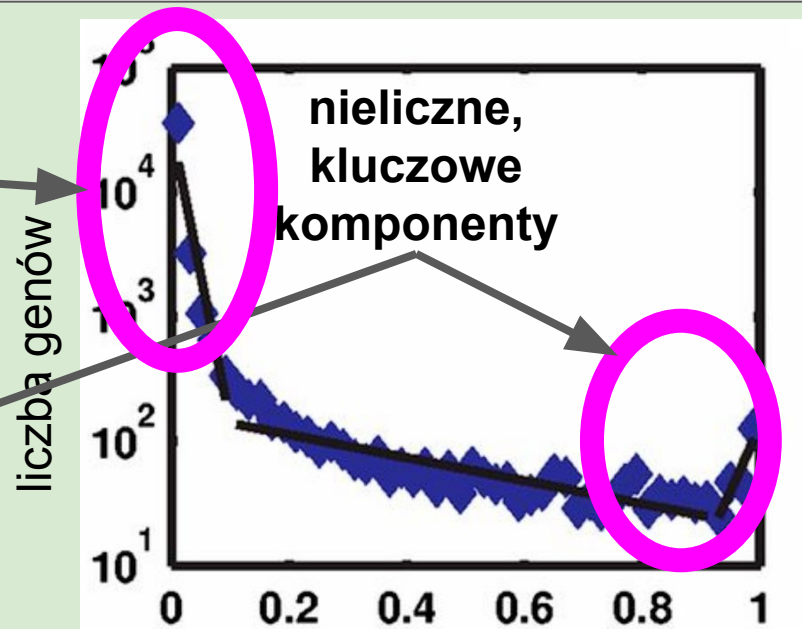


częstość występowania danego genu w genomach różnych gatunków bakterii

Wykorzystanie modułów



częstość instalacji danego pakietu na komputerach z systemem Linuks



częstość występowania danego genu w genomach różnych gatunków bakterii

Zagrożenia systemowe

Wirusy komputerowe

Powiera przez się przez skopiowanie do innych programów

Dokonuje zniszczeń po wystąpieniu pewnego zdarzenia

Może zmieniać kod aby uniknąć wykrycia przez program antywirusowy

Może blokować system antywirusowy

Wirusy biologiczne

Powiera się wewnątrz komórki wykorzystując jej zasoby

Replikują się w organizmie bez objawów

Mutacje sprawiają że stają się odporny na leki

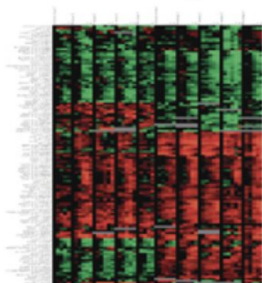
Może atakować system odpornościowy

Zmiana paradygmatu w naukach biomedycznych

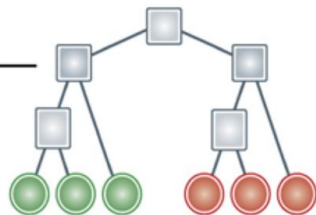
Tradycyjne podejście do prowadzenia badań

Biologia

Informatyka



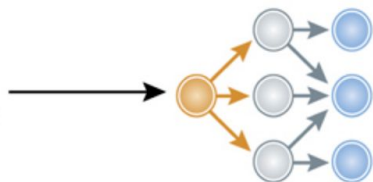
Mikromacierze



Grupowanie



Mózg ludzki

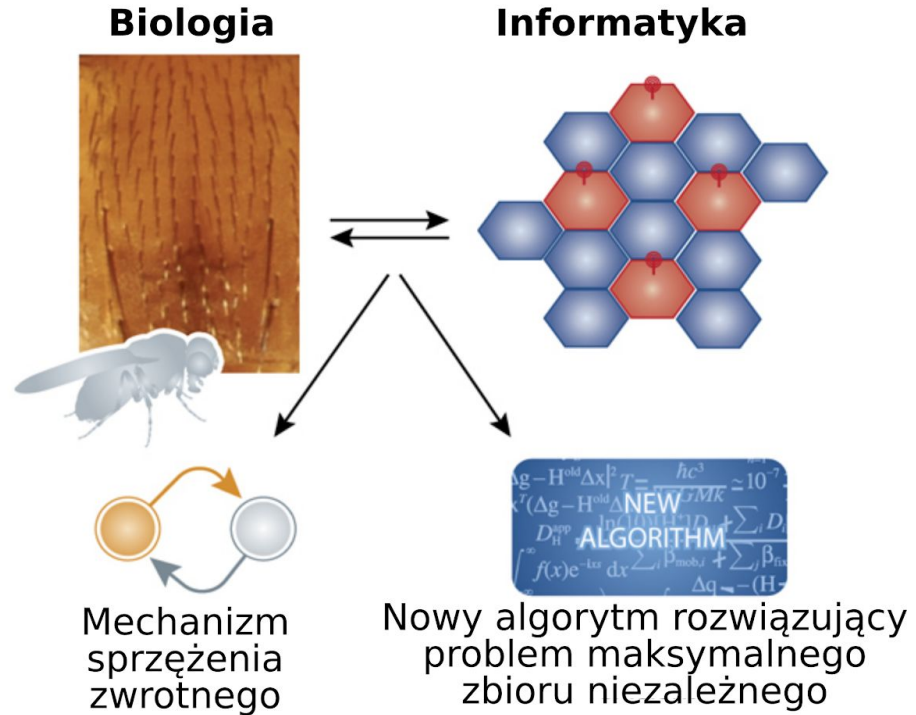
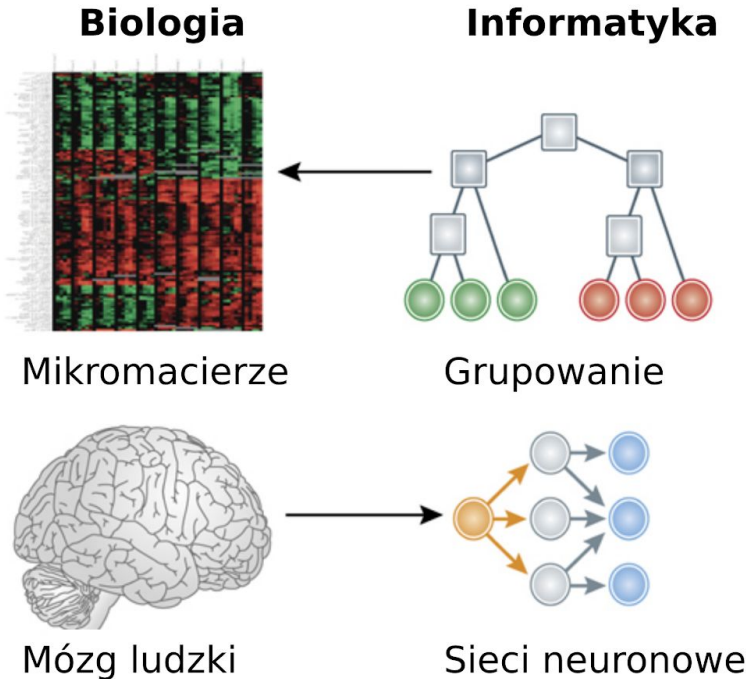


Sieci neuronowe

Zmiana paradygmatu w naukach biomedycznych

Tradycyjne podejście do prowadzenia badań

Myślenie obliczeniowe



Źródła: *Algorithms in nature: The convergence of systems biology and computational thinking.* S. Navlakha i Z. Bar-Joseph, *Molecular Systems Biology*, 2011. *A Biological solution to a fundamental distributed computing problem.* Y. Afek et al. *Science*, 2011

Zmiana paradygmatu w naukach biomedycznych

Dzięki odkryciom biologii molekularnej i genetyki w latach 60-tych biologia oraz biomedycyna stały się naukami o przetwarzaniu informacji.

Wprowadzono pojęcia: “**kod genetyczny**”, “**informacja genetyczna**”, “**transkrypcja**”, “**translacja**”.

Terminologie i koncepcje z informatyki i innych dyscyplin inżynierskich stają się powszechne w laboratoriach biologicznych i biomedycznych.

Kamienie milowe w rozwoju biomedycyny

- Wzrost mocy obliczeniowych, karty graficzne, FPGA
- Analiza sekwencji i algorytmy tekstowe
- Sztuczna inteligencja i algorytmy uczenia maszynowego
- Techniki Big Data i algorytmy rozproszone

- Odkrycie struktury DNA
- Projekt mapowania genomu ludzkiego
- Sekwencjonowanie Następnej Generacji
- Wielkoskalowe projekty sekwencjonowania
- Nowe narzędzia edycji genomowej (CRISPR-CAS9)

Kamienie milowe w rozwoju biomedycyny

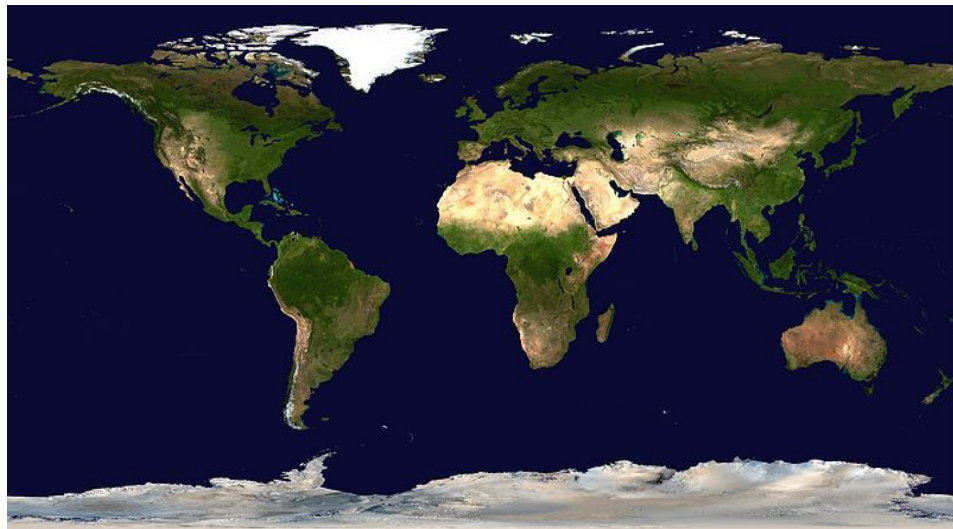
- Wzrost mocy obliczeniowych, karty graficzne, FPGA
- Analiza sekwencji i algorytmy tekstowe
- Sztuczna inteligencja i algorytmy uczenia maszynowego
- Techniki Big Data i algorytmy rozproszone

- Odkrycie struktury DNA
- Projekt mapowania genomu ludzkiego
- Sekwencjonowanie Następnej Generacji
- Wielkoskalowe projekty sekwencjonowania
- Nowe narzędzia edycji genomowej (CRISPR-CAS9)



Główne źródło sukcesów: Zarówno dane jak i algorytmy są udostępnione publicznie, na zasadach otwartych licencji

Ewolucja w pozyskiwaniu danych

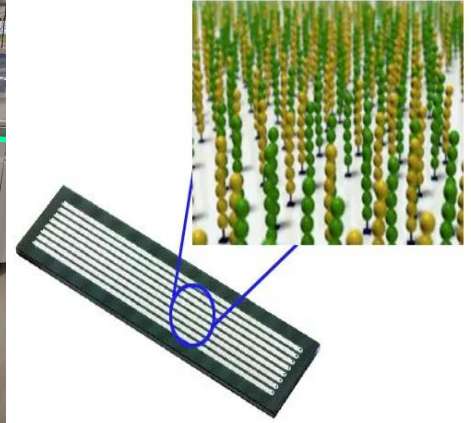
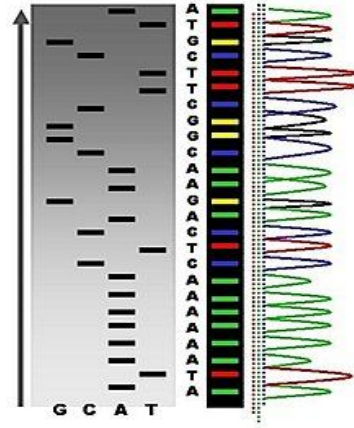


- "Teatr świata" - 1570 r.
- Opracowywano z lądu i morza
- Tworzono przez kilkadziesiąt lat

- Obraz satelitarny
- Generowany w ciągu 1 sekundy
- Bardziej informatywny i dokładny

Źródła: https://en.wikipedia.org/wiki/Theatrum_Orbis_Terrarum,
https://en.wikipedia.org/wiki/File:Whole_world_-_land_and_oceans_12000.jpg

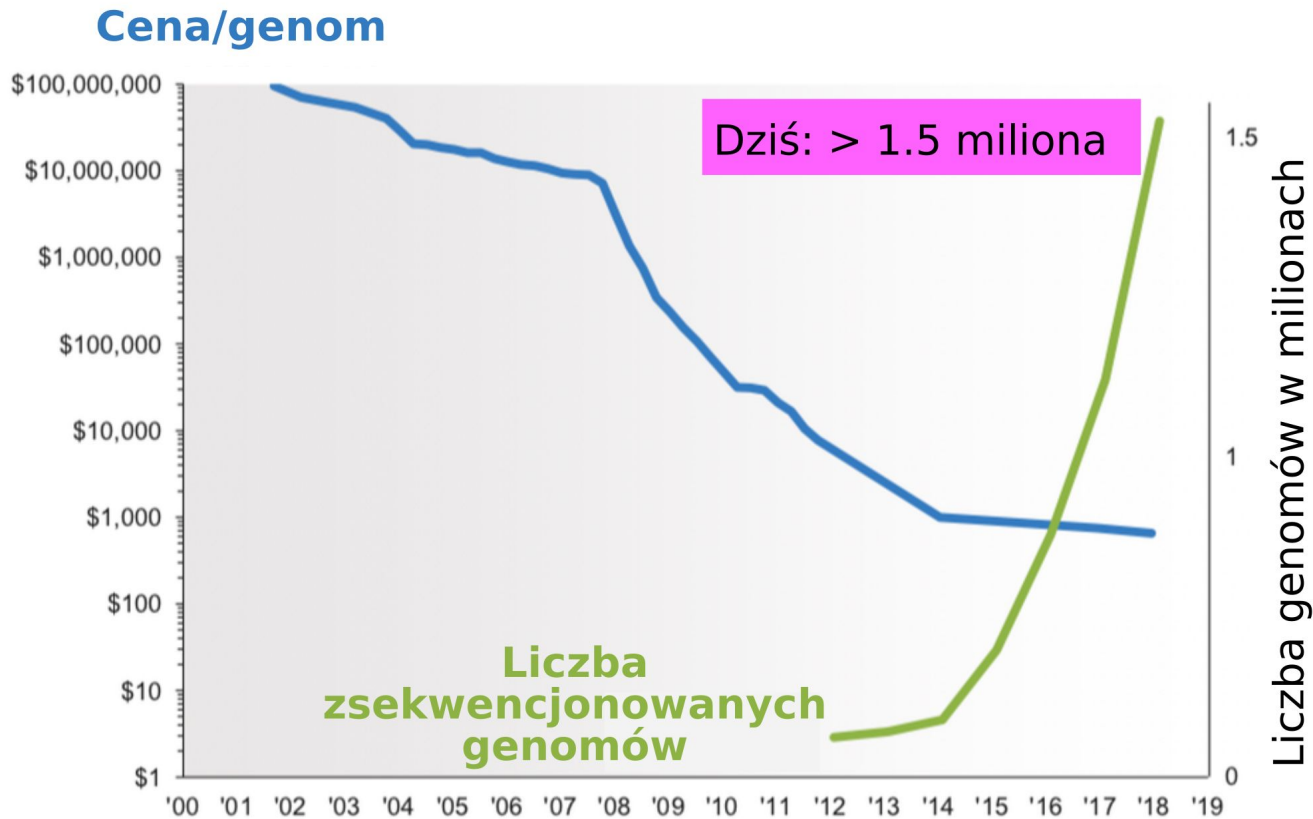
Ewolucja w pozyskiwaniu danych



- **PCR, elektroforeza na żelu i sekwencjonowanie metodą Sangera**
- Analiza krótkich fragmentów genomu, pojedynczych genów
- Czasochłonne przygotowanie próbek

- **Sekwencjonowanie Następnej Generacji**
- Jednoczesne sekwencjonowane **miliardów** krótkich fragmentów DNA
- Szybka analiza całych genomów
- Robotyzacja procesu przygotowania próbki


Genomiczne Big Data



200 mln - 2 mld
genomów ludzkich do
2025 roku

2 - 40 Exabajtów (EB)
1 EB = 1 mln TB

Roczne
zapotrzebowanie
YouTube w 2025
~ **1-2 EB**



**Diagnostyka molekularna
(choroby rzadkie)**

**Epidemiologia,
wirusologia**

**Projektowanie
leków**

Bioinformatyka

Metagenomika

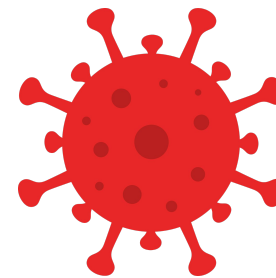
Terapie genowe

Medycyna prewencyjna

Medycyna personalizowana

Bioinformatyka w dobie SARS-CoV2

- Diagnostyka
- Epidemiologia - modelowanie statystyczne
- Analiza genomu wirusa
- Projektowanie leków, szczepionek
- Analiza genomu pacjenta
- Śledzenie wirusów w przestrzeni publicznej (Metasub)



Źródło: Małopolskie Centrum Biotechnologii



Autor: CRG, Christine Panagiotidis

A portrait of Eric Lander, a man with grey curly hair and a mustache, wearing a blue shirt and a dark jacket, smiling against a brown background.

Time for the FUTURE!!

*We will see young scientists in the future
writing D++ (DNA Code)*

Eric Lander MIT and Harvard



Dziękuję!

Zapraszam do kontaktu: team@biodatageeks.org