

Python w bioinformatyce

Bartek Wilczyński

bartek@mimuw.edu.pl

<http://regulomics.mimuw.edu.pl/>

Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski



Pystok #37

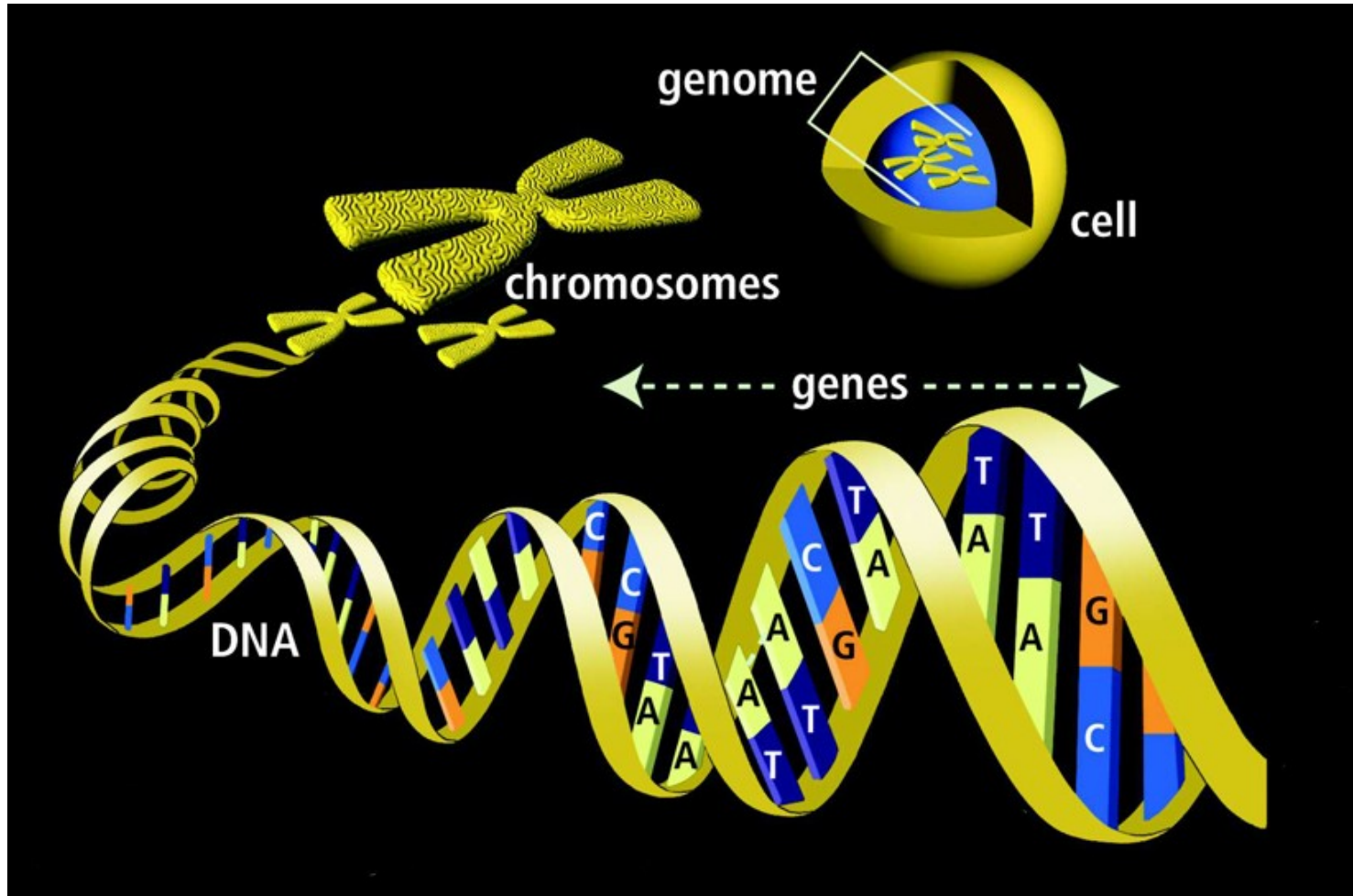


16. stycznia 2019

**Bioinformatics is a
discipline devoted to
studying of information
processes in biotic systems**

P. Hogeweg (1978)

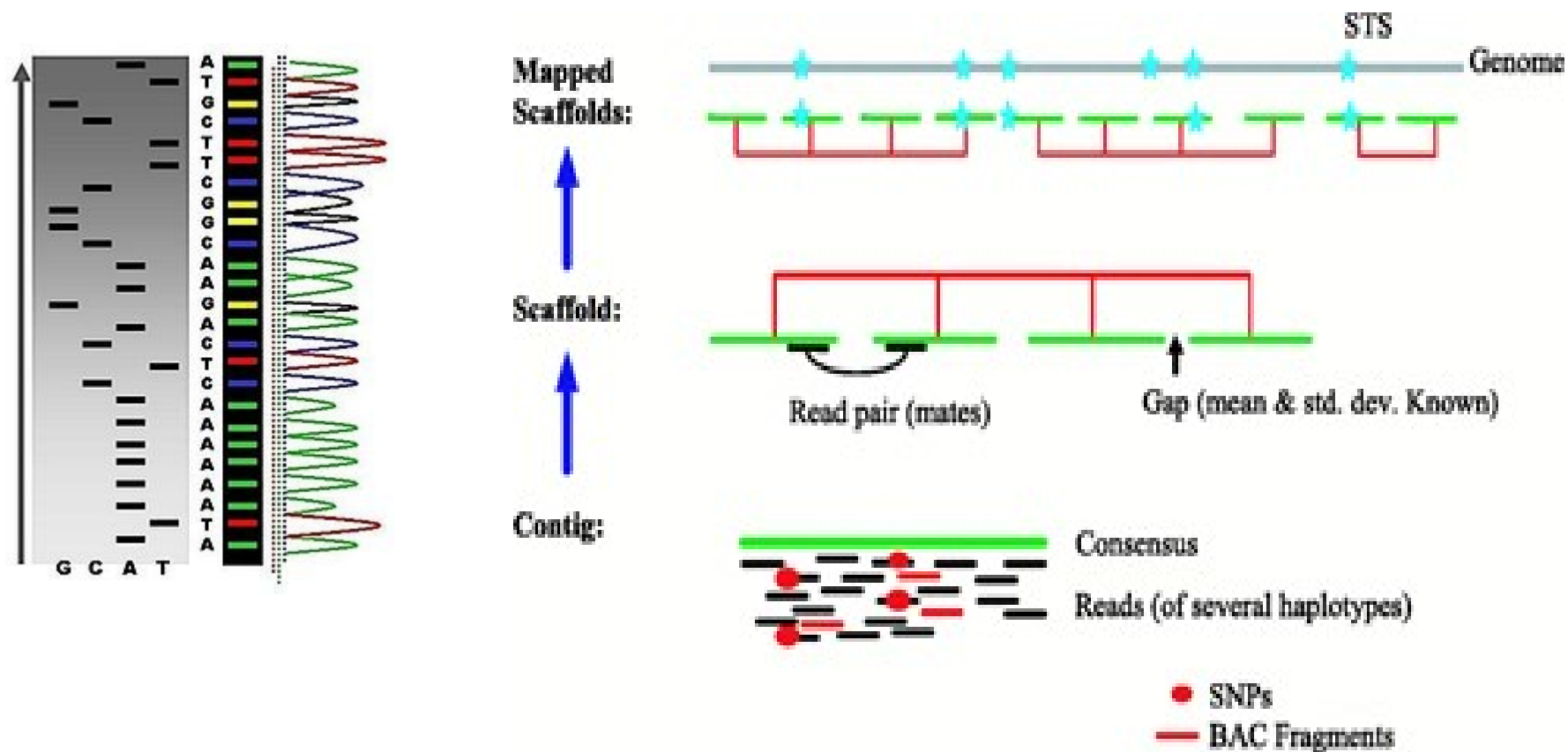
Informacja w DNA



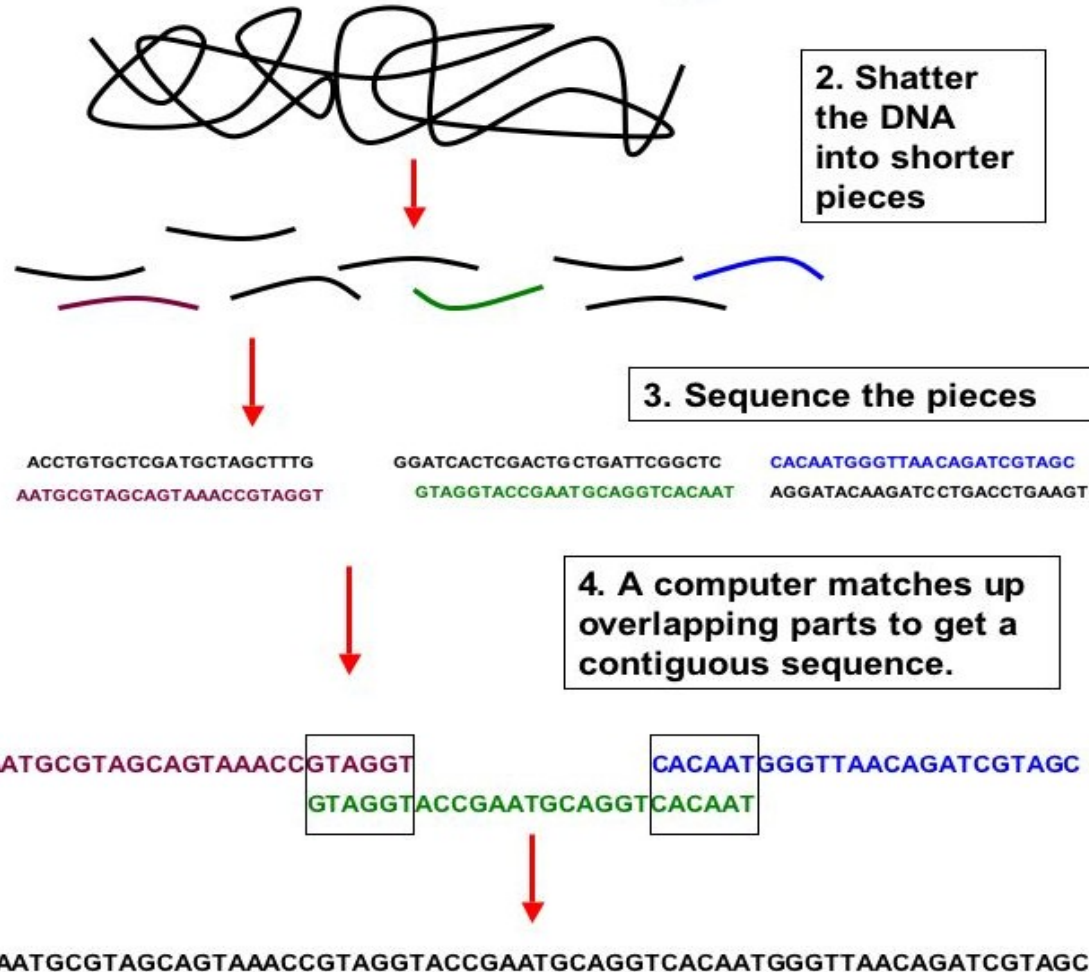
Sekwencjonowanie DNA

- **Human Genome Project**
 - Ogromny projekt rządu USA
 - 3 miliardy dolarów
 - 15 lat badań
- **Celera genome project**
 - Konkurencyjny projekt komercyjny
 - 300 milionów dolarów
 - 2-3 lata badań

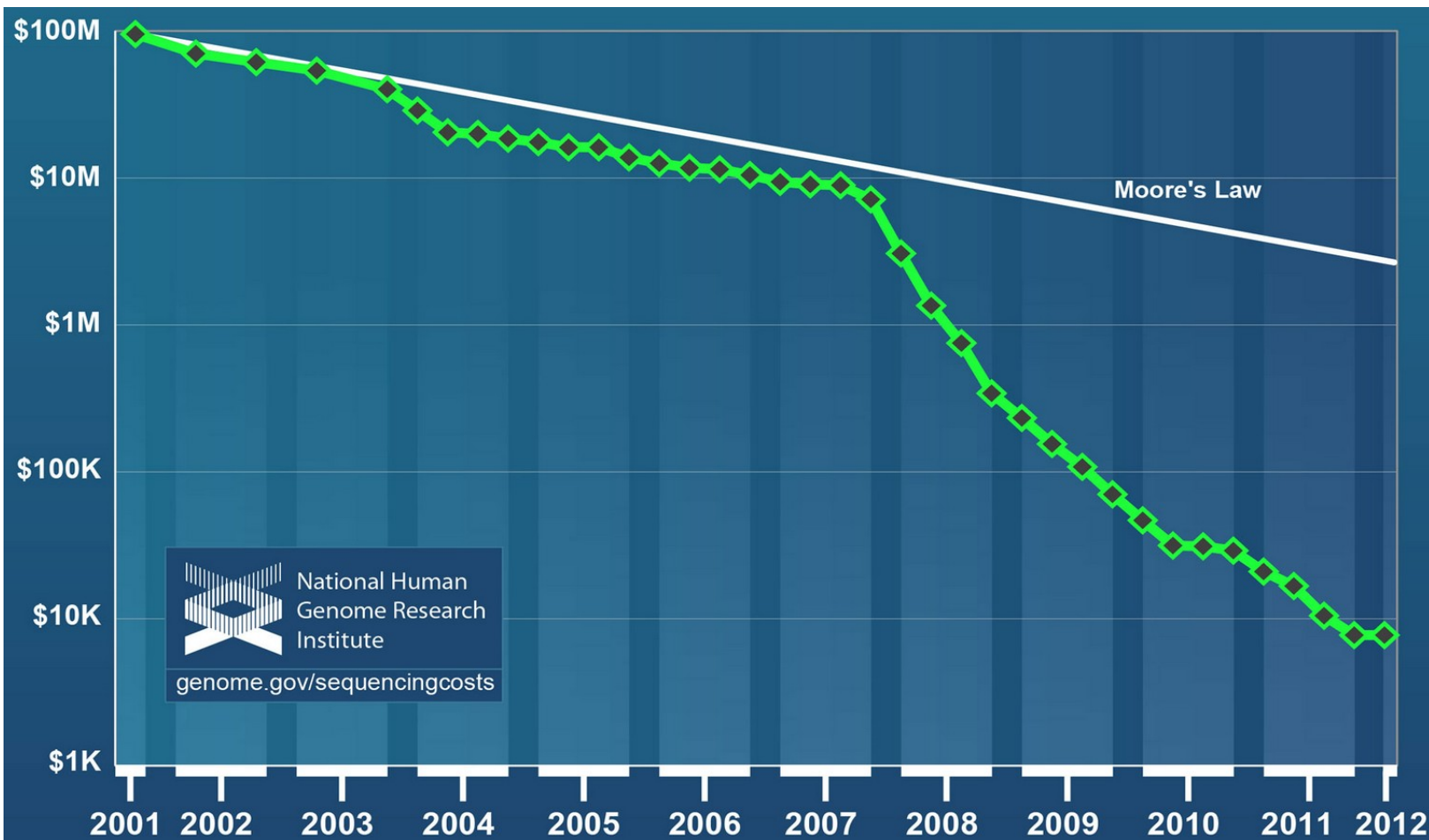
Sekwencjonowanie DNA



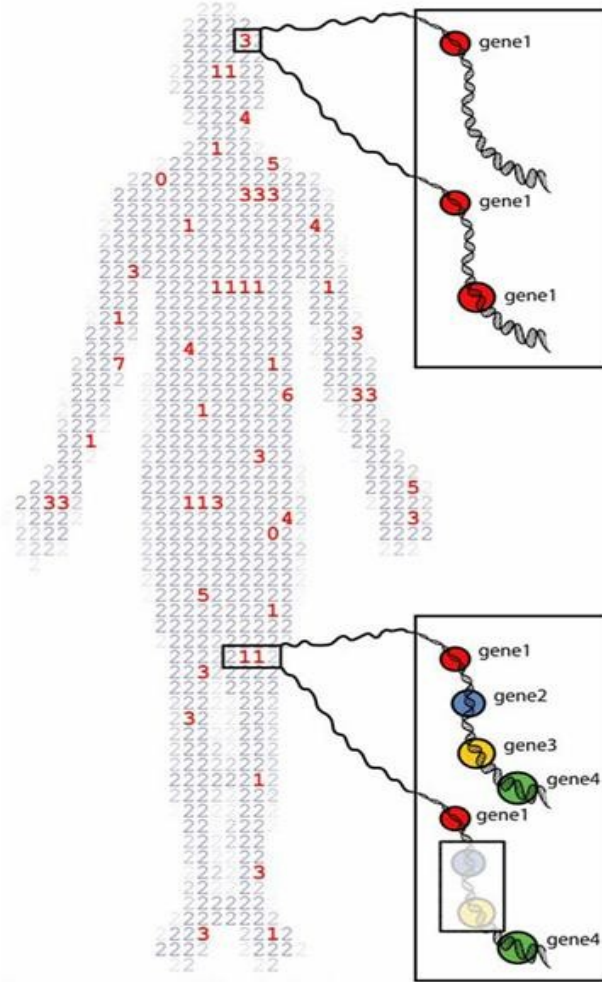
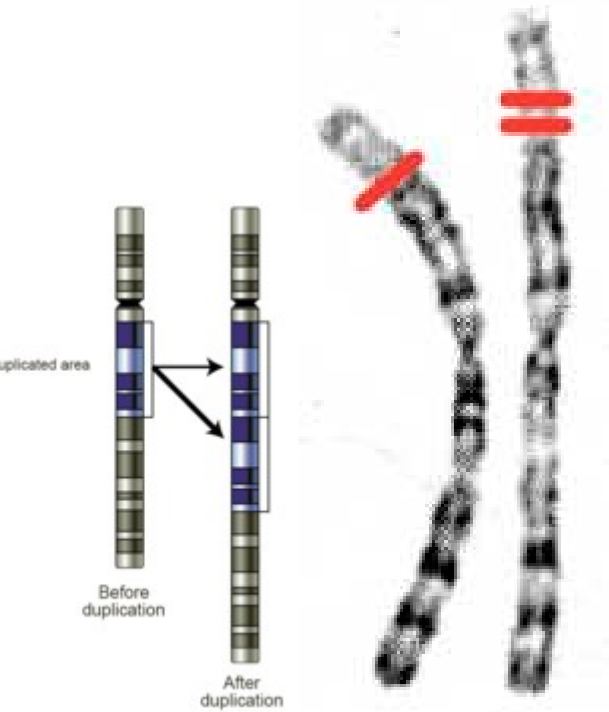
Sekwencjonowanie DNA



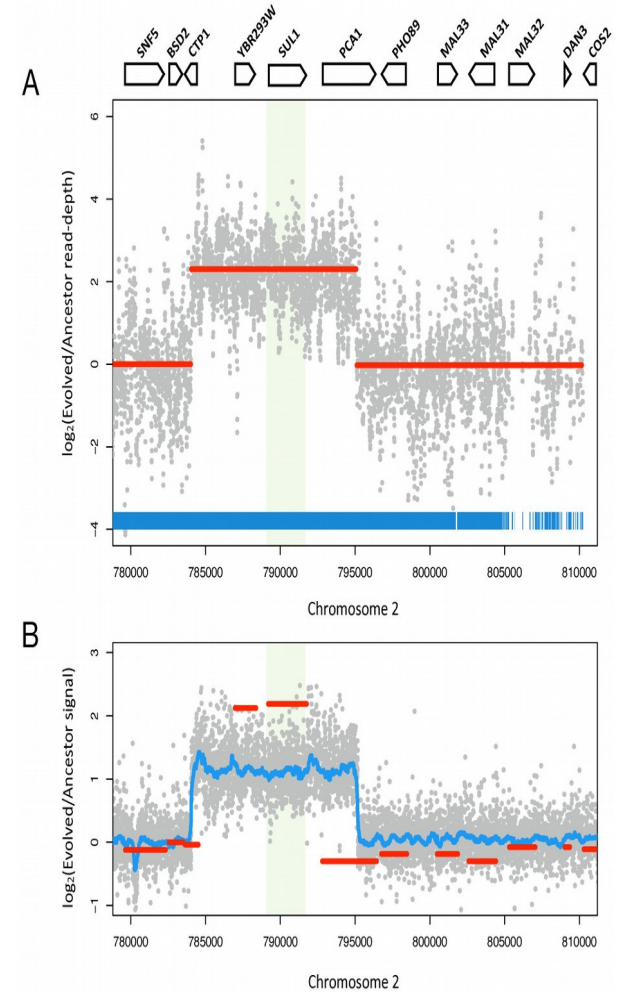
Sekwencjonowanie DNA



Mutacje interesują pacjentów



Copy number variation in the human genome



Komercjalizacja genomu



[HOME](#) [COMPANY](#) [SCIENCE](#) [PUBLICATIONS](#) [NEWS](#) [CAREERS](#) [CONTACT](#)



Are you a medical professional? [Get More Info](#) | [Enroll Practice](#) | [Order](#)
Already have a kit? [Activate Kit](#) | [Log In](#) | [Physician Portal](#)

[HOME](#) [COMPANY](#)



[Главная](#) [Услуги](#)



+7 (495) 767-85-95

[搜索](#)

[遗传课堂](#) | [科学新闻](#) | [博客](#) | [帮助](#)

[立即拨打](#) [4008-731-120](#)

The largest DNA ancestry service in the world

[sign in](#)

[register kit](#)



[welcome](#)

[ancestry](#)

[how it works](#)

[buy](#)

[search](#)

[help](#)

welcome
to you

Find out what your
DNA says about you
and your family.

Trace your lineage back 10,000 years and discover your
history from over 750 maternal lineages and over 500
paternal lineages.

Krótką historia biopythona

- 1999 – Jeff Chang i Andrew Dalke tworzą biopython'a na podstawie bioperl'a
- 2002 - pierwsze wydanie
- 2003-4 – Brad Chapman kieruje pracami
- 2005- Peter Cock przejmuje kierownictwo
- 2009 – publikacja
- 2010 – przeniesienie rozwoju na git/github
setki ludzi, miliony linii kodu, tysiące funkcji,
dziesiątki tysięcy użytkowników

biopython / biopython

Unwatch 143 Star 1,369 Fork 759

Code Issues 245 Pull requests 63 Insights

Pulse

Contributors

Community

Traffic

Commits

Code frequency

Dependency graph

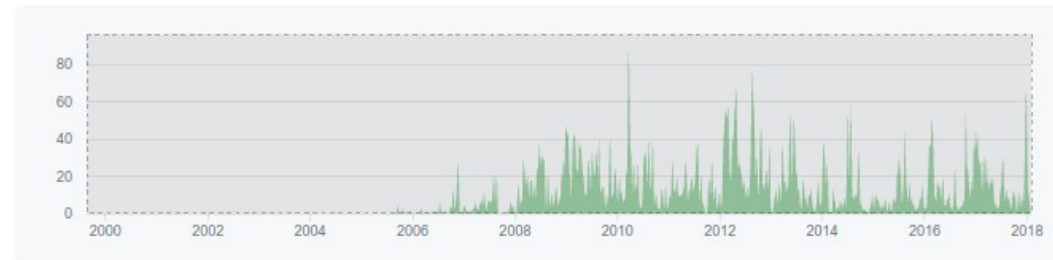
Network

Forks

Dec 5, 1999 – May 18, 2018

Contributions: Commits

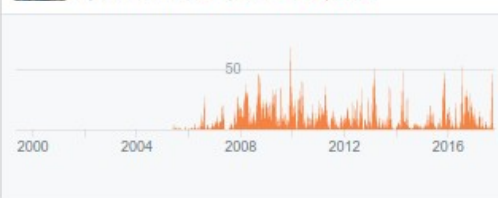
Contributions to master, excluding merge commits



peterjc

#1

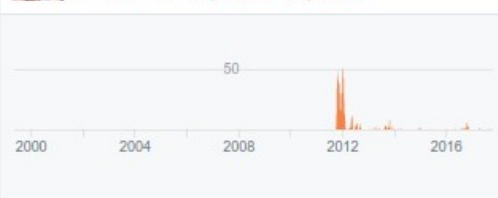
6,128 commits 471,907 ++ 406,249 --



bow

#2

711 commits 196,450 ++ 131,255 --



etal

#3

513 commits 46,752 ++ 23,630 --



cbrueffer

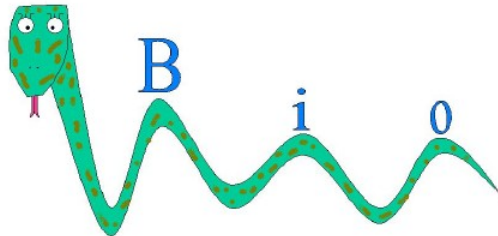
#4

163 commits 15,701 ++ 13,205 --

Promocja biopython'a w 2003 :)

Open-Bio + Python = Biopython

- Official blurb: international association of developers of freely available Python tools for computational molecular biology; established in 1999.
- <http://biopython.org>
- Library of functionality for dealing with common problems biologists programming in python might face.



Logo w roku 2003 :) ->

© Brad Chapman

Repositories 10K

Code 1M

Commits 39K

Issues 8K

Marketplace

Topics 24

Wikis 3K

Users 3K

Languages

| | |
|------------------|-------|
| Python | 3,249 |
| R | 764 |
| Java | 560 |
| Jupyter Notebook | 543 |
| HTML | 526 |
| Perl | 469 |
| C++ | 431 |
| Shell | 399 |
| JavaScript | 299 |
| C | 204 |

[Advanced search](#) [Cheat sheet](#)

10,780 repository results

Sort: Best match

Languages

| | |
|------------------|-------|
| Python | 3,249 |
| R | 764 |
| Java | 560 |
| Jupyter Notebook | 543 |
| HTML | 526 |
| Perl | 469 |
| C++ | 431 |
| Shell | 399 |
| JavaScript | 299 |
| C | 204 |

★ 355

★ 852

ile ★ 270

★ 703

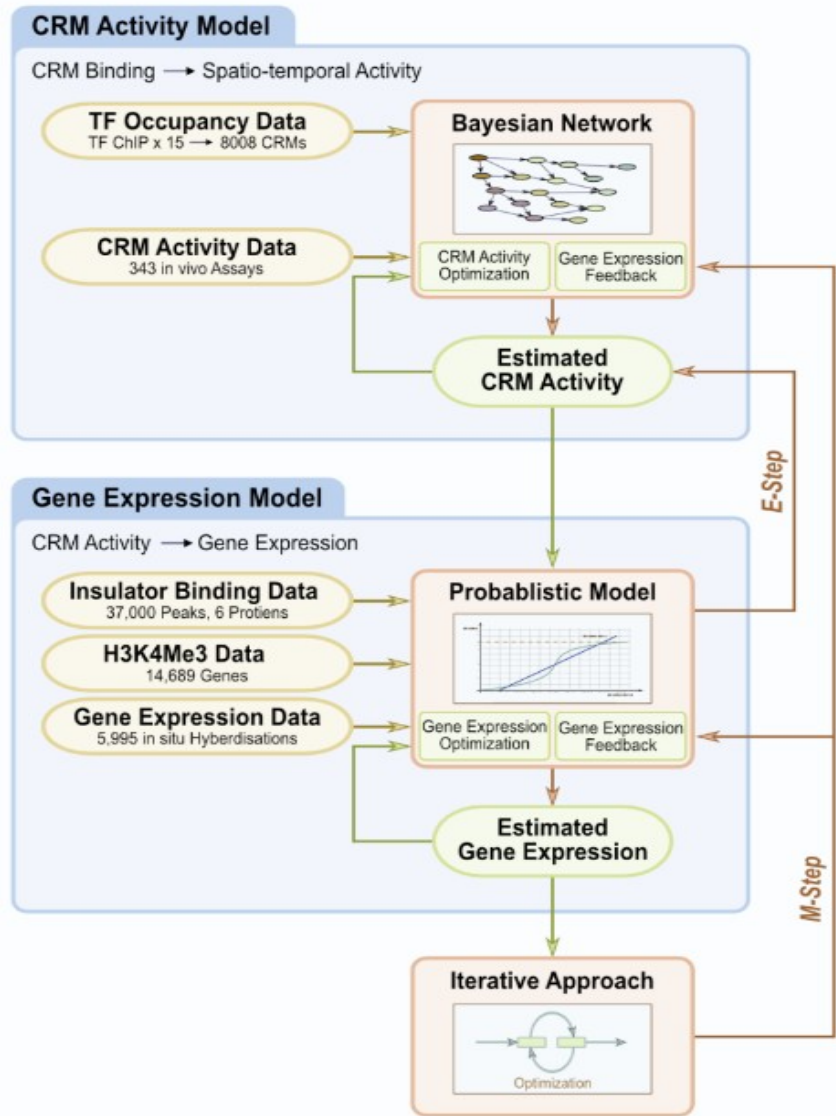
★ 284

Co można w biopythonie

- I/O dla wielu formatów plików,
 - Porównywanie sekwencji i struktur,
 - Podstawowe algorytmy bioinformatyczne
-
- Ja np. Prowadzę cały semestralny kurs “Wstęp do Biologii obliczeniowej” dla Informatyków

Przykładowy projekt

- Kilka tysięcy genów
- Kilka tkanek
- Kilka faz rozwoju zarodkowego
- Kilkanaście zmiennych powiązanych
- **Kto jest aktywny kiedy, gdzie i dlaczego?**



- 8008 enhancers compiled from 15 ChIP experiments (almost 20k binding peaks)
- Activity data for ~140 enhancers divided into
 - 3 tissues (MESO, VM, SM)
 - 5 stages (4-6, 7-8, 9-10, 11-12, 13-16)
- Gene expression data for 5082 genes from the BDGP database

```
def learn(self, genes, g_exp, dist, binding, crm_act, promoters):
```

```
    crm_bind=binding.select(ids=crm_act.elems)  
    f=crm_bind.to_bnf(crm_act.get_column(crm_act.header[0]))  
    d=data.dataset().fromNewFile(f)
```

```
    sc,g,subp=d.learn(score="BDE",data_factor=DATA_FACT,prior=None,verbose=True,n_min=NET_LIM,
```

```
    self.cpd=d.to_cpd(g.weighted_edges(subp,MIN_EDGE_W))
```

```
def predict(self, testing, dist_test, binding, promoters):
```

```
    bind_test=binding.select(ids=crms_test)  
    f=bind_test.to_bnf()  
    d=data.dataset().fromNewFile(f)
```

```
    crm_class=d.classify_list(self.cpd,prob=1)
```

```
    res=[]
```

```
    for g in testing:
```

```
        r=1.0
```

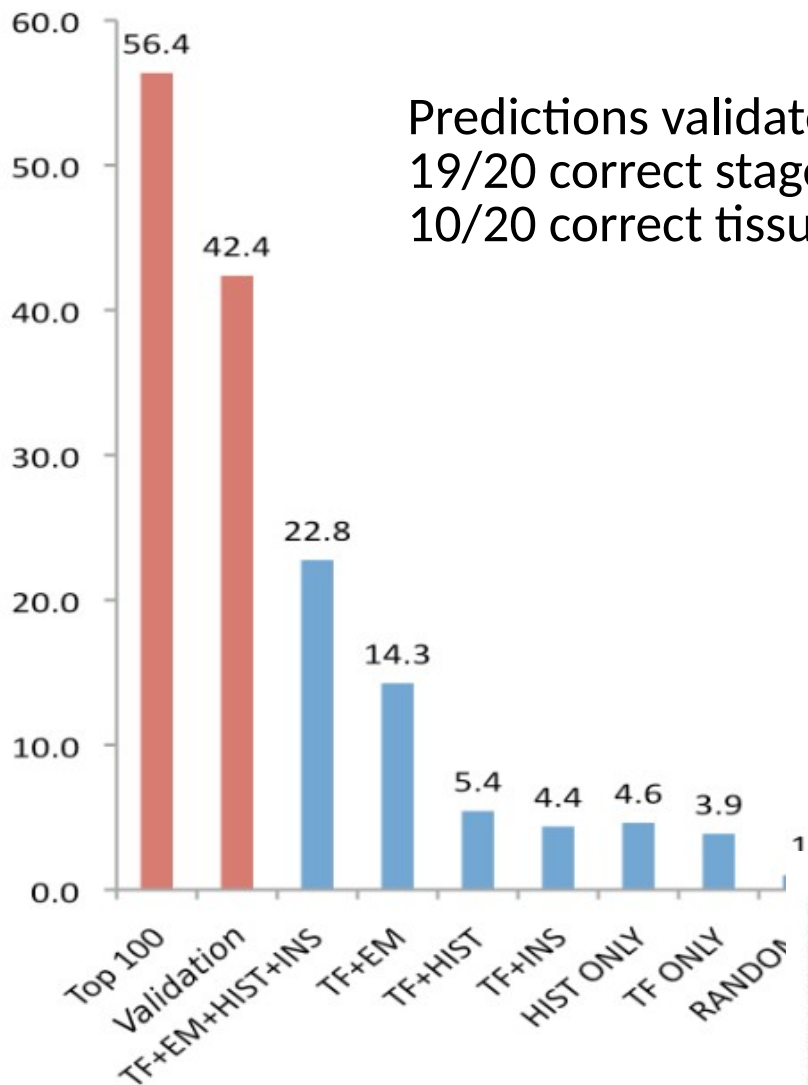
```
        for c,dist in dist_test[g]:
```

```
            r*=(1-crm_class[c,"Class"])
```

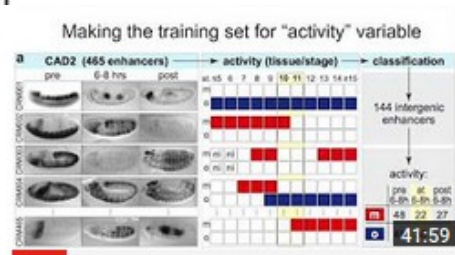
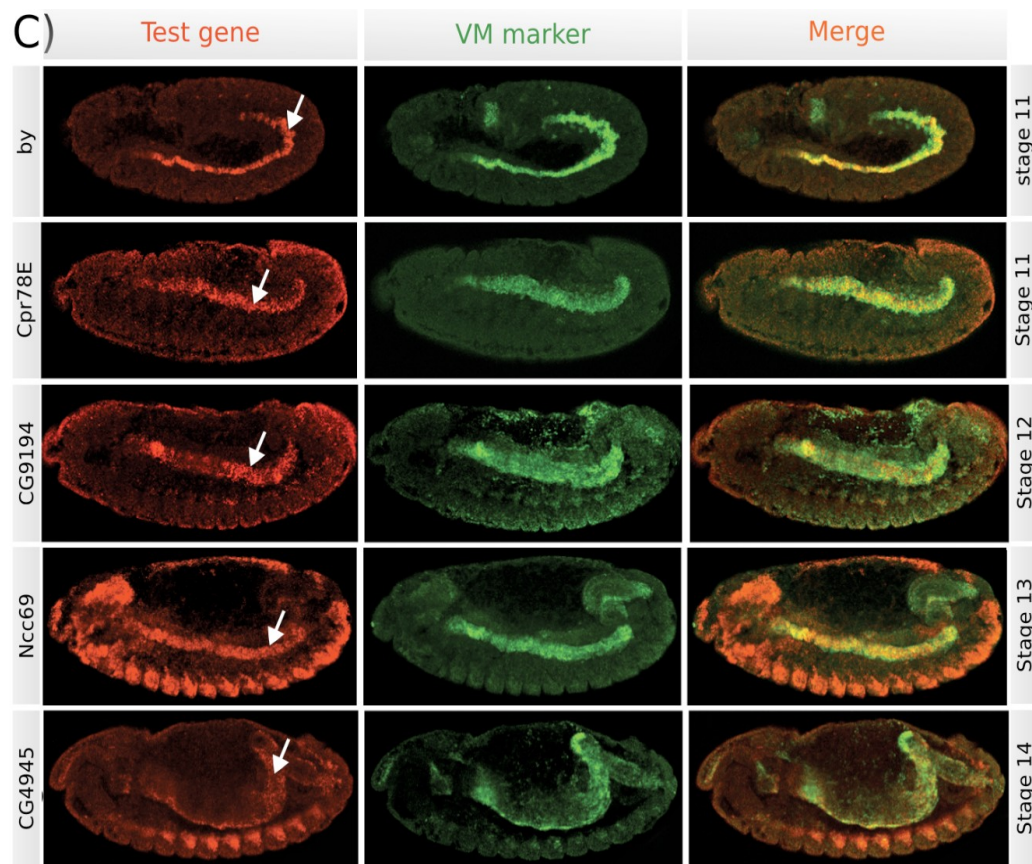
```
        res.append((1.0-r)*promoters[g])
```

```
    return res
```

Enrichment of correct predictions



Predictions validated:
19/20 correct stage,
10/20 correct tissue



Bartek Wilczynski - Using Python to Find a Bayesian Network Describing Your Data

PyData

2 years ago • 5,670 views

PyData SV 2014 Today's world is full of data that is easily accessible for anyone. The problem now is how to make sense of this ...

<http://youtu.be/WD7b1jLeZZ4>

Typowy zestaw bibliotek w projekcie bioinformatycznym
(w mojej grupie)

- * **biopython** – basic I/O
- * **pandas** – reprezentacja danych tablicowych
- * **scikit-learn** (albo BNfinder) – uczenie maszynowe
- * **matplotlib/seaborn** – wizualizacja danych

OpenSource – czy warto ?

- Zawsze pisząc program można się czegoś nauczyć...
- Jeśli udostępnimy to innym, to może ktoś to przeczyta i coś nam podpowie...
- Jeśli go to zainteresuje, może dołączy do projektu...
- Jeśli projekt okaże się sukcesem, może stać się ważnym dla szerszej społeczności...

Czasem warto mądrze przyłączyć się, zamiast robić wszystko od nowa

Leadership Lessons from Dancing Guy: The First Follower and | TED-Ed - Chromium

Leadership Lessons x

Secure | https://ed.ted.com/on/lgslePtt


Antek

Leadership Lessons from Dancing Guy: The First Follower and more...

LESSON CREATED BY **AMANDA BJORN** USING **TED Ed**
VIDEO FROM **Derek Sivers** YOUTUBE CHANNEL

Let's Begin...

Today, we are going to be learning some lessons about leadership from a shirtless dancing guy. We will watch dancing guy create a movement, from start to finish, in under 3 minutes, and during this, we will hear from Derek Sivers as he dissects some fascinating lessons about leadership.



Watch

Think

Dig Deeper

...And Finally

Customize This Lesson

78

Create and share a new lesson based on this one.

- Paweł Bednarz
- Ilona Grabowicz
- Julia Herman-Iżycka
- Magdalena Machnicka
- Ania Macioszek
- Marlena Osipowicz
- Agnieszka Podsiadło
- Irina Tuszyńska
- Rafał Zaborowski

- Eileen Furlong (EMBL)
- Mara Dierssen (CRG)
- Minna Kaikonnen (UEF)
- Carla Margulies (LMU)



NARODOWE CENTRUM NAUKI



EUROPEAN UNION
EUROPEAN REGIONAL
DEVELOPMENT FUND

