

Mining and Processing Biomedical Data

Dr. rer. nat. Krisztian Buza

adiunkt naukowy

Faculty of Mathematics, Informatics and Mechanics

University of Warsaw, Poland

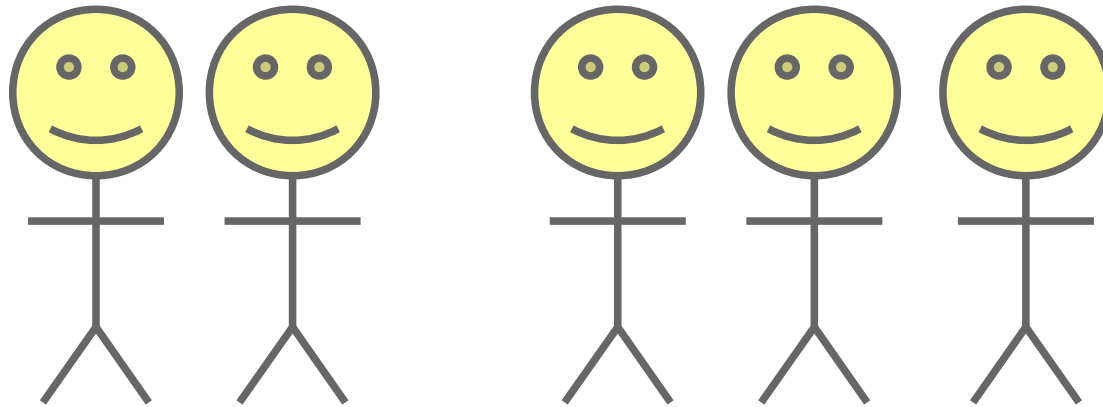
chrisbuza@yahoo.com

Revolution in Biology

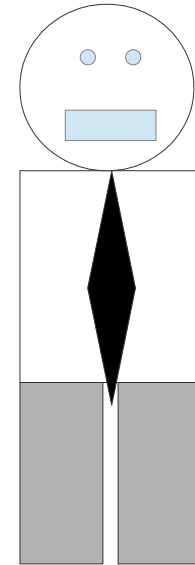
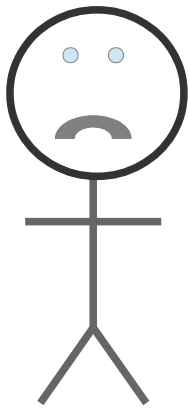
An overview of some of the most surprising recent results

We will be able to heal cancer better!

But this technology may have bad applications...



Sorry, you have
bad genes



Example: human height

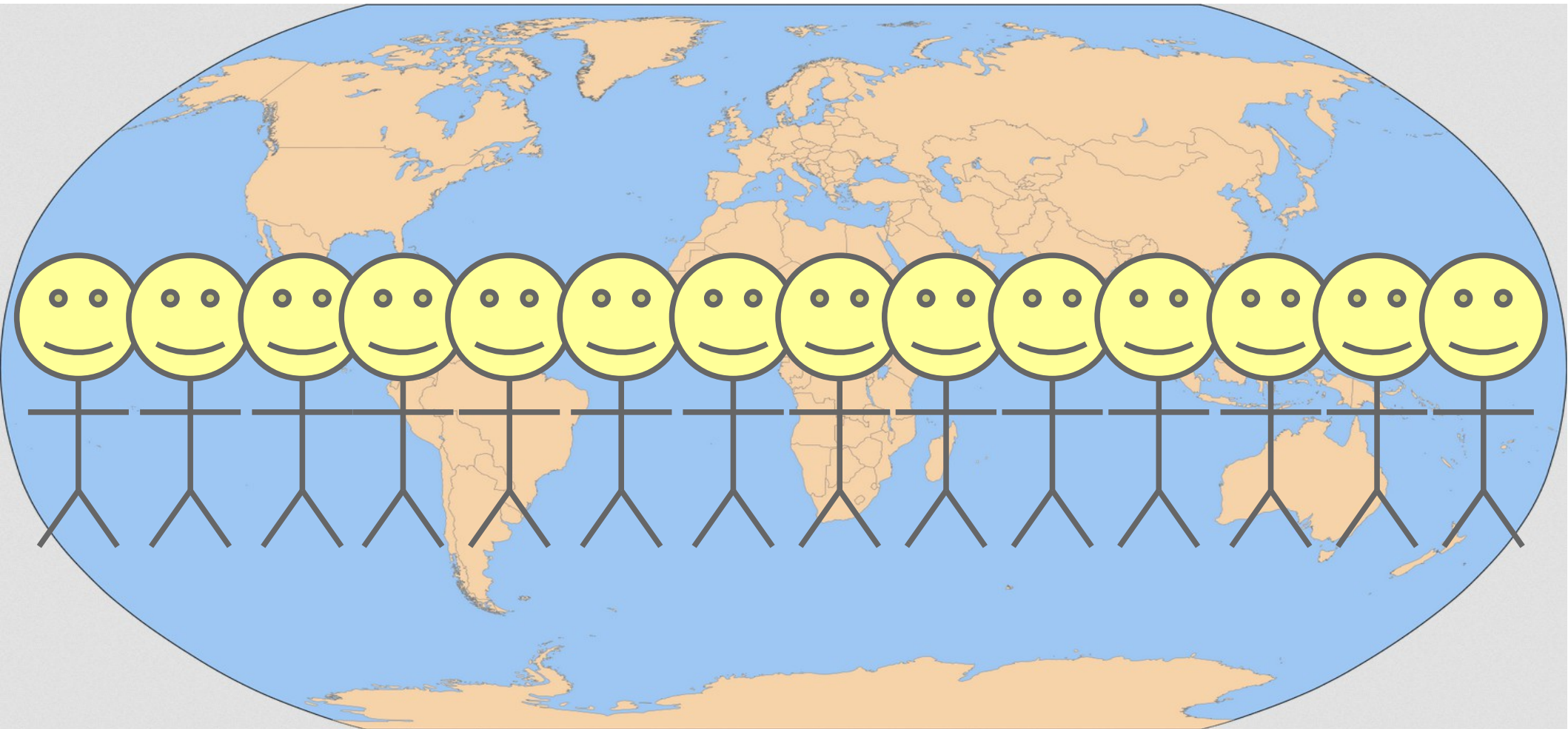
- **80%** of the observed differences in human height can be attributed to the differences in the DNA sequence between human individuals
- Researchers identified approx. 40 genes
- All these 40 genes together, explain only approx. **5%** of the observed differences

Publication of the human genome sequence

- What we know
 - Sequence of DNA of the “average” human
 - Most common variations
- What is just being discovered
 - The associations between phenotype and DNA variants

We all have a lot in common...

DNA sequences of two **unrelated humans** have a similarity of **99.9%**

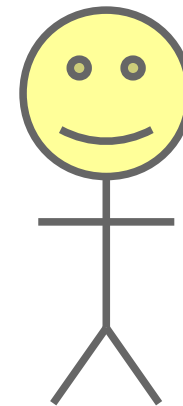


A public-domain image from Wikipedia is used on this slide



...but

DNA sequences of a **monkey** and a **human** have a similarity of **98.5%**



from
Zahīr ud-Dīn Muḥammad Bābur
(1483-1530): Bāburnāma

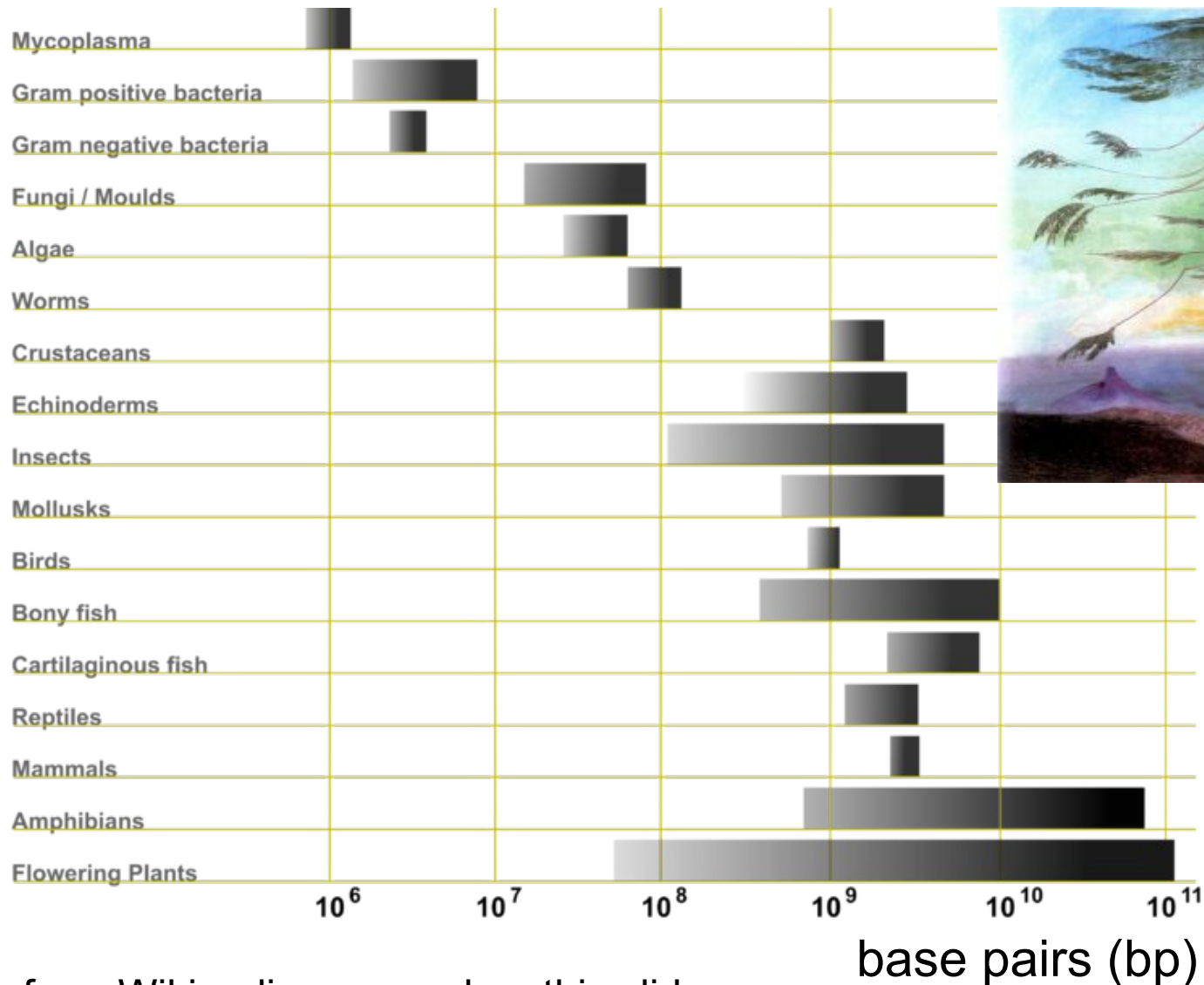
“Neandertals and Moderns made imperfect mates”,
Science (31 January 2014), Vol. 343, pp. 471 - 472



Image from Wikipedia. Licence info:

http://en.wikipedia.org/wiki/File:Homo_neanderthalensis_adult_male_-_head_model_-_Smithsonian_Museum_of_Natural_History_-_2012-05-17.jpg

Genome sizes



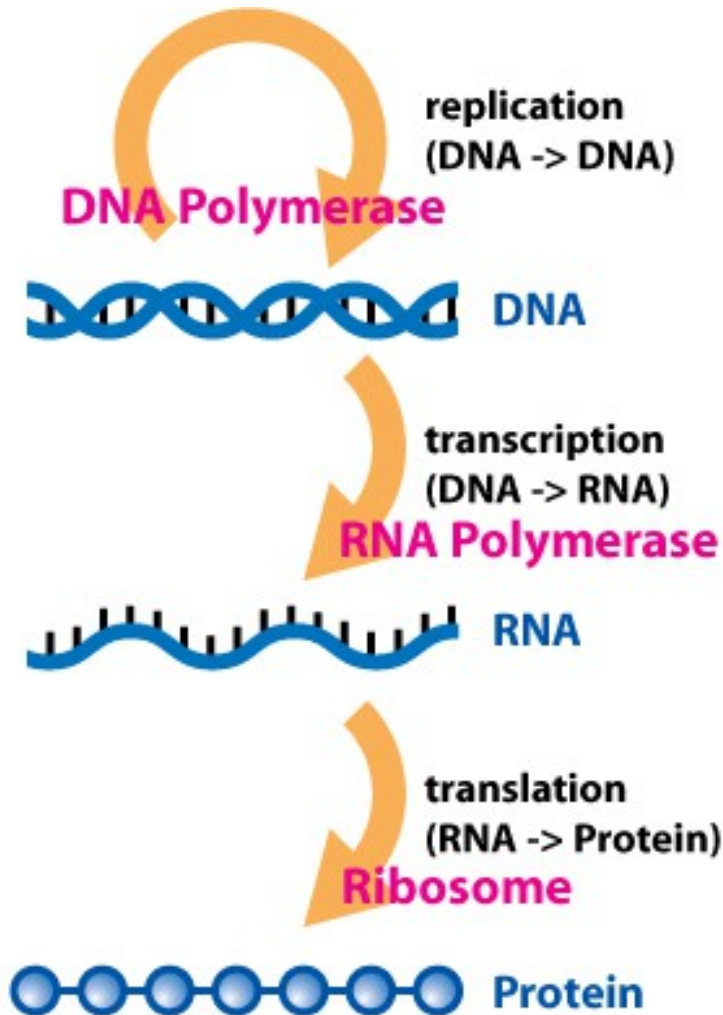
Csontváry K. Tivadar:
Magányos cédrus
("Lonely cedar")

Images from Wikipedia are used on this slide.

Licence info: http://en.wikipedia.org/wiki/File:Genome_Sizes.png

http://en.wikipedia.org/wiki/File:Cskt-maganyos_cedrus_%281907%29.jpg

Transcription and Translation of the DNA



Only ~1– 2% of the human DNA codes for proteins

Gene: a segment of the DNA that codes for a protein

(only protein-coding genes are considered in this talk)

~23 000 genes in humans

Image from Wikipedia. Licence info:

http://en.wikipedia.org/wiki/File:Central_Dogma_of_Molecular_Biochemistry_with_Enzymes.jpg

Mutation of the “Fruitless” gene



Drosophila / Fruit fly

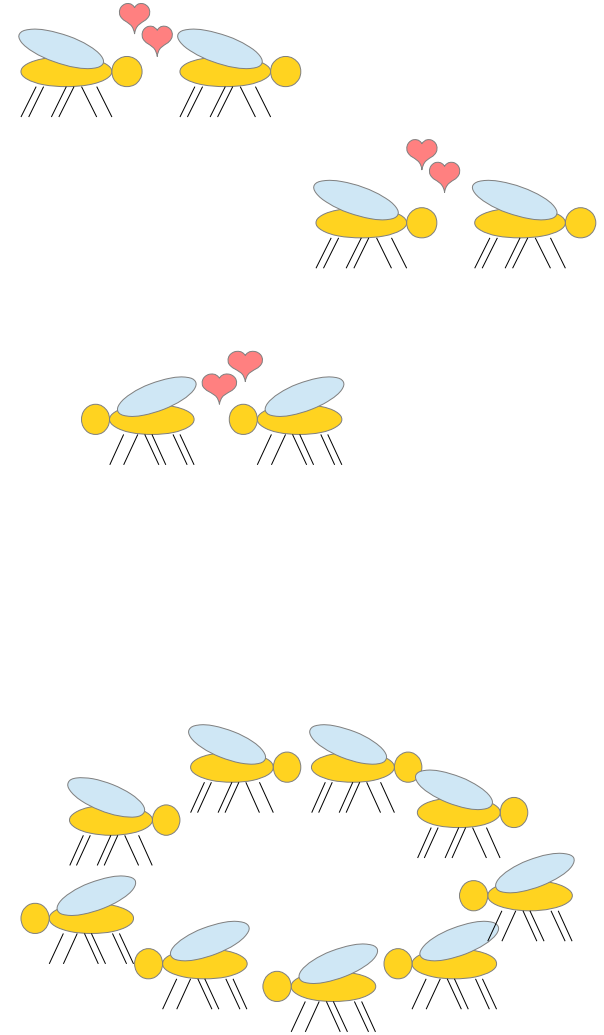
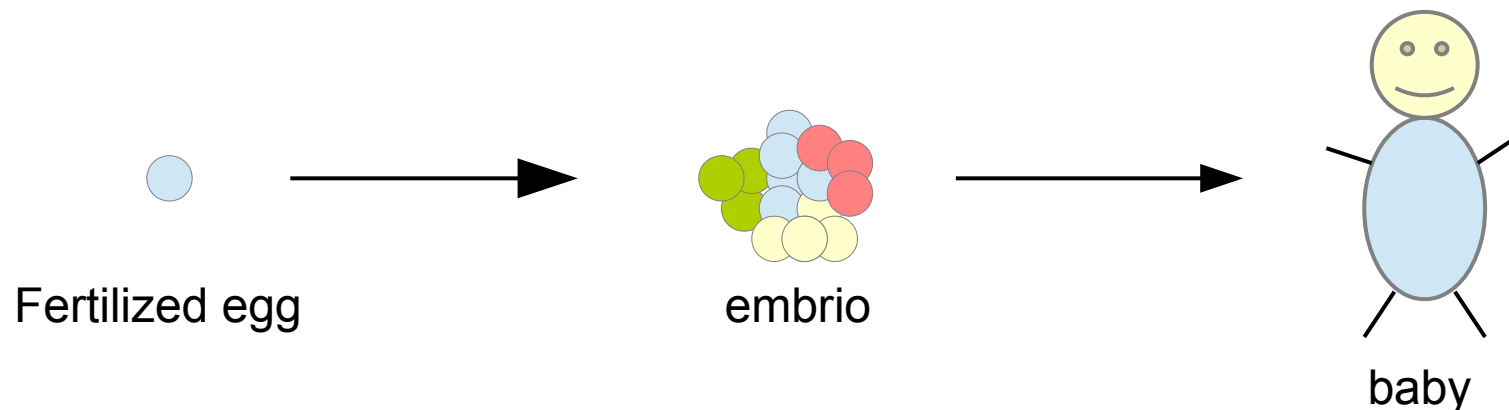


Image from Wikipedia. Licence info:
http://en.wikipedia.org/wiki/File:Drosophila_repleta_lateral.jpg

Gene Expression

- DNA → RNA → protein
- Genes can be turned “on” / ”off”:
 - turned on: expressed, turned off: silenced
 - gene expression levels (real numbers)
- Environmental factors may influence gene expression (see e.g. multifactorial diseases)



An extreme example of gene expression: X-inactivation



Optional homework: please read
http://en.wikipedia.org/wiki/Calico_cat