

Genomics and Bioinformatics

Introduction on *evo/eco/cons* inferences from genomic data using R

Matteo Fumagalli

Using genomics to infer the evolutionary history of a species

History + Genomics = ???

History + Genomics = ???



Donald J. Trump [@realDonaldTrump](#)



Sorry losers and haters, but my I.Q. is one of the highest -and you all know it! Please don't feel so stupid or insecure,it's not your fault

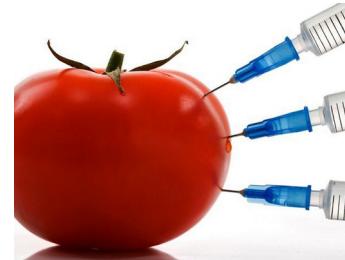
[Reply](#) [Retweet](#) [Favorite](#) [More](#)

1,105 RETWEETS 863 FAVORITES



6.37 PM - 8 May 13

History + Genomics = ???



Donald J. Trump [@realDonaldTrump](#)



Sorry losers and haters, but my I.Q. is one of the highest -and you all know it! Please don't feel so stupid or insecure,it's not your fault

[Reply](#) [3 Retweet](#) [★ Favorite](#) [More](#)

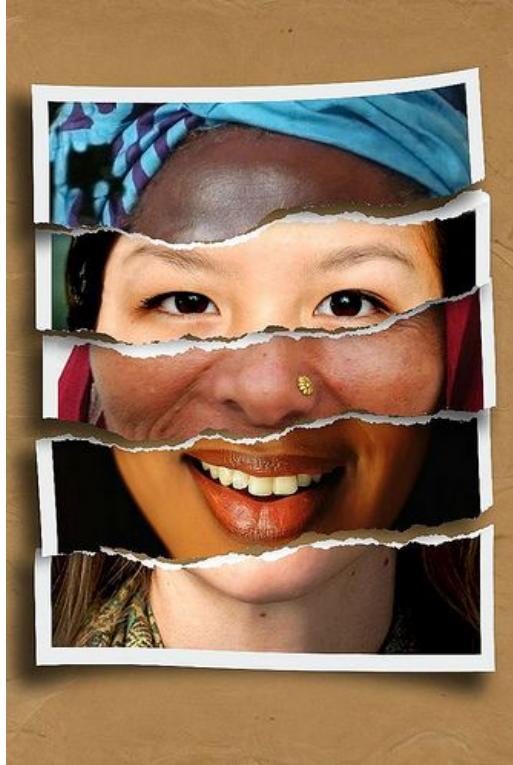
1,105 RETWEETS 863 FAVORITES

6.37 PM - 8 May 13

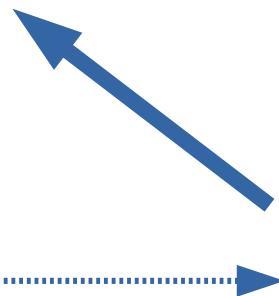


How can we use genomics to learn something about the history of a species (or population)?

Everyone of us is unique



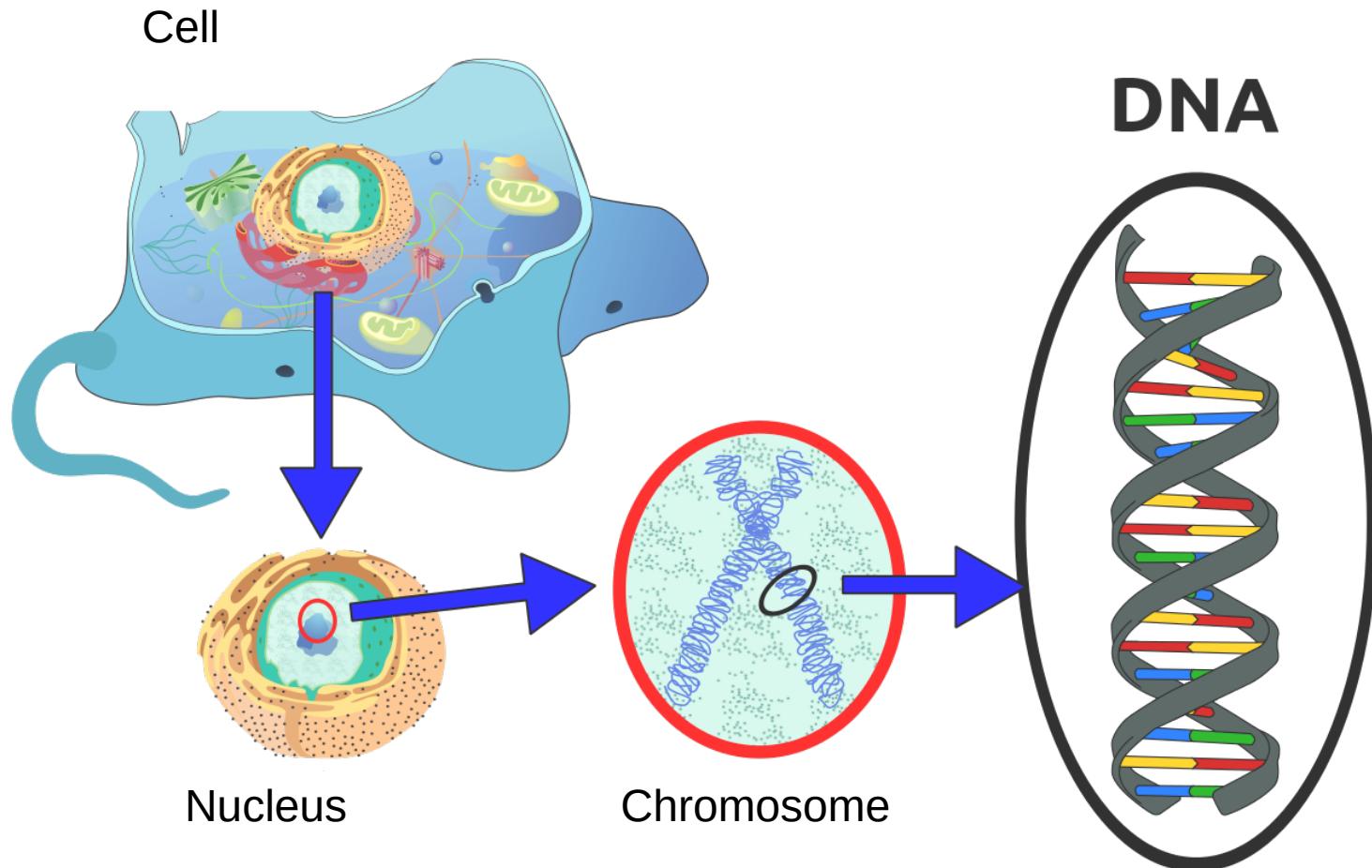
You



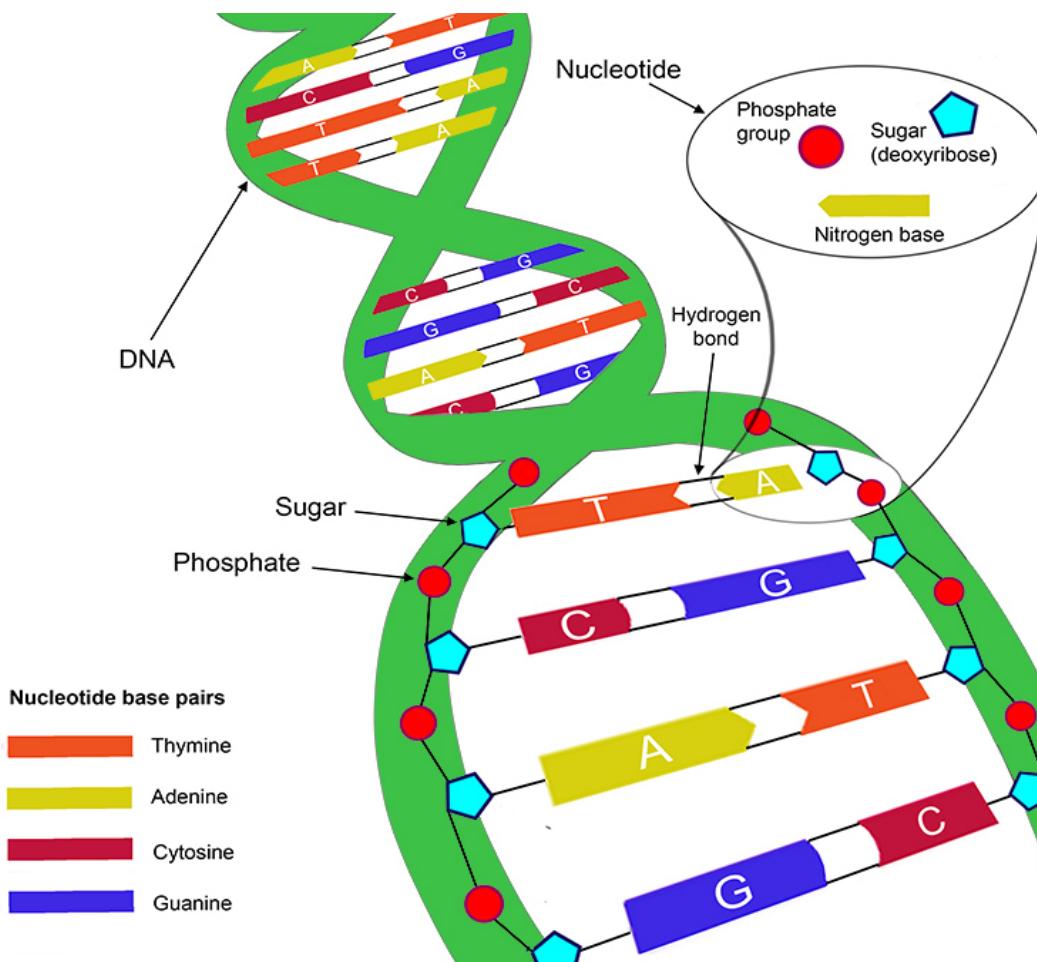
Environment



Nucleus of a human cell hosts 23 pairs of chromosomes which contain the DNA, our genetic “luggage” (genome).



A paired sequence of 4 possible nucleotides: A, C, G and T.



DNA (genome)

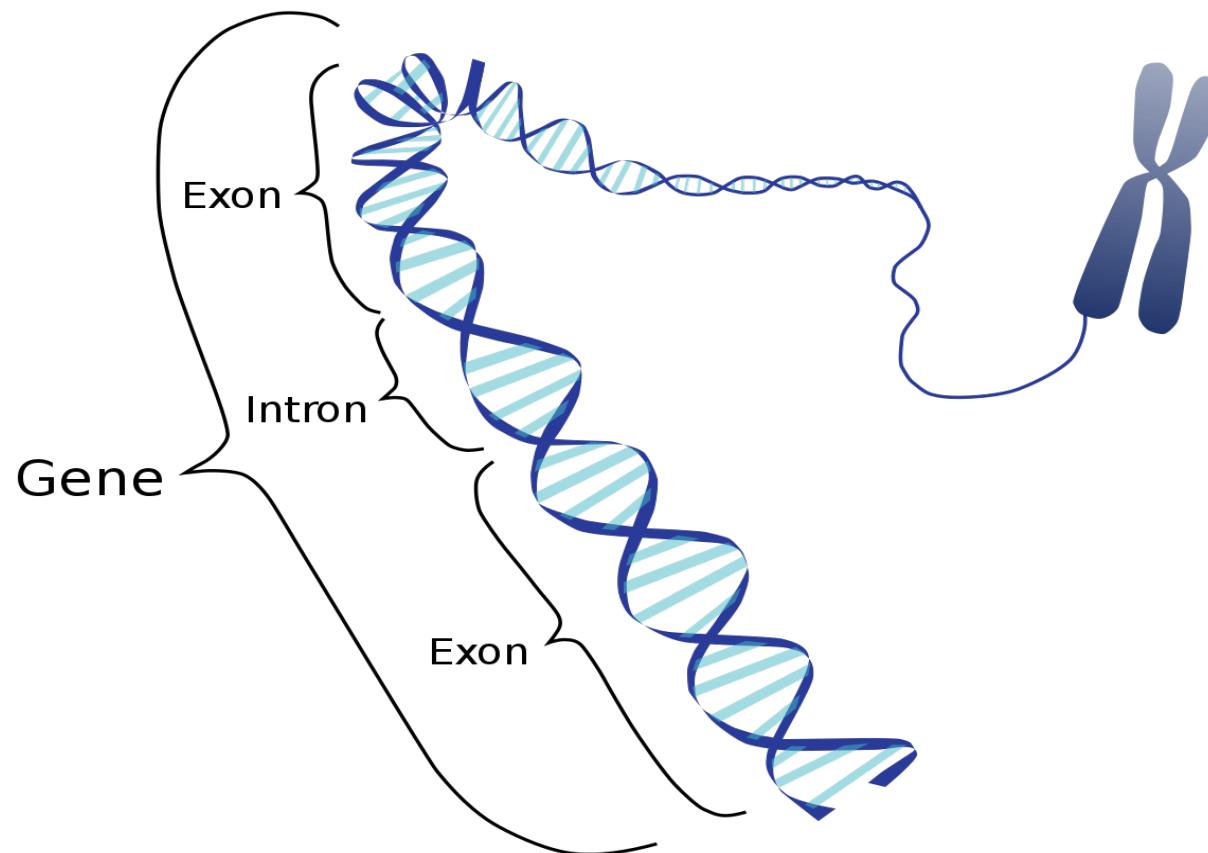
AACGCTGAGATCGCGAAATGGCGCTCGAGAGGCATAAGCGATCGAGCGAGCATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATA
TGCAGCGAGCTATCGATCGAGCGAGGCATTATTTCGAGGCAGGGCGAGCTATTATGGCGAGCGAGGCAGCGATATTATTAA
TTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATGCATCGTAGCTAGTGATCGCTAGCTAGATGTCACTGAG
CTAGCATTATATACTGCGGCTAGTCAGTCGGCGCGCGATATTACGTAGCTAGCGATCGTAGCTAGCTAGCGATCGTAGCTAG
TCAGTACGCTAGTCATGCTAGCTAGCGATGCATGCTAGCTAGCTAGCGATCGATGTATGCTATGCGATGCGTAGCTATTAGTCGATCGCGATGCTAG
CGATGCGATGCGTAGTCGATGCTAGCTAGCTAGCGATCGATGTATGCTATGCGATGCGTAGCTATTAGTCGATCGCGATGCTAG
CTAGTCGATGCTAGCTAGCATGCTAGATGTTAGTCGCGGATTATGCTAGCTAGTCGATGCTACGTAGCTAGTCGACGCTAG
TAGCTGATCGCTATGCTATCGAGCTAGTCGCTAGCTATGCATCGTAGCTATGCGTAGCTGGCTAACGCTGAGATCGCGAAATGGCGCT
CGAGAGGCTAACGCGATCGAGCGAGCATCGAGCAGAGCGAGGGCGATCTATGGCTATATTAAATATGCGAGCGAGCTATCGATCGAGCGA
GGCGATTATTATTGAGGCAGGGCGAGCTATTATGGCGAGCGAGGCAGCGAGCGATATTATTATTATATCGGGCGCGCGCAGCG
CAGAGACGCGAATTATCGAGCGATGCATCGTAGCTAGTGATCGCTAGCTAGCGATGCTAGCTAGCGATGCGTAGTCGATGCTAG
GTCGGCGCGCGCGCGATATTACGTAGCTAGCGATGCTAGCTAGCGATCTAGCGATGCTAGCTAGCGATGCGTAGTCGATGCTAG
ATCGATGCATGCTAGCTAGCTAGCGATGCTAGCTAGCGATGCTAGCTAGCGATGCTAGCTAGCGATGCGTAGTCGATGCTAG
CTAGCTAGCTATCGATCGATGTATGCTATGCGTAGCTATTAGTCGATCGCGATGCTAGCTAGTCGATGCTAGCTAGCGAT
CTAGATGTTATCTAGCTCGCGATTATGCTAGCTAGTCGATGCTACGCTAGCTAGCTGCTATGCTATCGAGCTA
GTCAGTCGCTAGCTATGCATCGTAGCTATGCGTAGCTGGCTAACGCTGAGATCGCGAAATGGCGCTCGAGAGGCTAACGCGATCGAGCGAG
CATCGAGCAGAGCGAGGGCGATCTATGGCTATATTAAATATGCGAGCGAGCTATCGATCGAGCGAGGGCGATTATTATTGAGGCAGGGCG
AGCTATTATGGCGAGCGAGGCAGGCAGCGAGCGATATTATTATTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATG
CATCGTAGCTAGTGATCGGCTAGCTAGATGTCAGTCAGCTAGCATTATATAGTCGCGCTAGTCAGTCGCGCAGCGAGCGAGGGCGATAT
TACGTAGCTAGCGAGCTAGCGATCGTAGCTAGCTAGTCAGTCAGTCAGTCGATGCTAGCTAGCGATGCGTAGTCGATGCTAGCTAG
AGCTAGTCGATCGATGCTAGCGATCTAGCGATGCTAGCTATCGATCGCGATGCGTAGTCGATGCTAGCTAGCTAGCTATCGATCGATGTATGC
TATGTCATCGTAGCTATTAGTCGATCGCGCGATGCTAGCTAGTCGATGCTAGCTAGCTAGCATGCTAGCTAGCTAGCTAGTTATCTAGCTCGCGGATTA
TGTCTAGCTAGTCGATGCTACGTTAGTCGACGCTAGCTAGCTGATCGCTATGCTATCGAGCTAGTCGCTAGCTATGCTATCGCTAT
GCTATGCGTAGGGCTAACGCTGAGATCGCGAAATGGCGCTCGAGAGGCTAACGCGATCGAGCGAGCGATCGAGCAGAGCGAGGGCGATCT
ATCGGCTATATTAAATATGCGAGCGAGCTATCGATCGAGCGAGGCAGGGCGATTATTATTGAGGCAGGGCGAGCTATTATGGCGAGCGAGGCAG
GCGAGCGATATTATTATTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTGATCGGCTAG
CTAGATGTCAGTCAGCTAGCATTATATAGTCGCGCTAGTCAGTCGCGCGCGCGAGCTATTACGTAGCTAGTCGAGCTAGCTAG
CGATCGTAGCTAGCTAGCGTAGTCGATGCTAGCTGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCTAGTCGATGCGTAGCTAG
TCTAGCGATGTAGCTATCGATCGCGATGCGTAGTCGATGCTAGCTAGCTAGCGATGCTAGCTAGCTAGCGATGCTAGTGCTATGCGTAGCTATTAGTC
GATCGCGCGATGCTAGCTAGTCGATGCTAGCTAGCGATGCTAGCTAGCGATGCTAGCTAGTCGATGCGTAGCTAGTCGATGCTACGTA
GCTTAGTCGACGCTAGCTAGCTGATCGCTATCGAGCTAGTCGAGCTAGTCGATGCGTAGCTAGCGATGCGTAGCTAGTCGATGCGTAGCTAGCGA

DNA (genome)

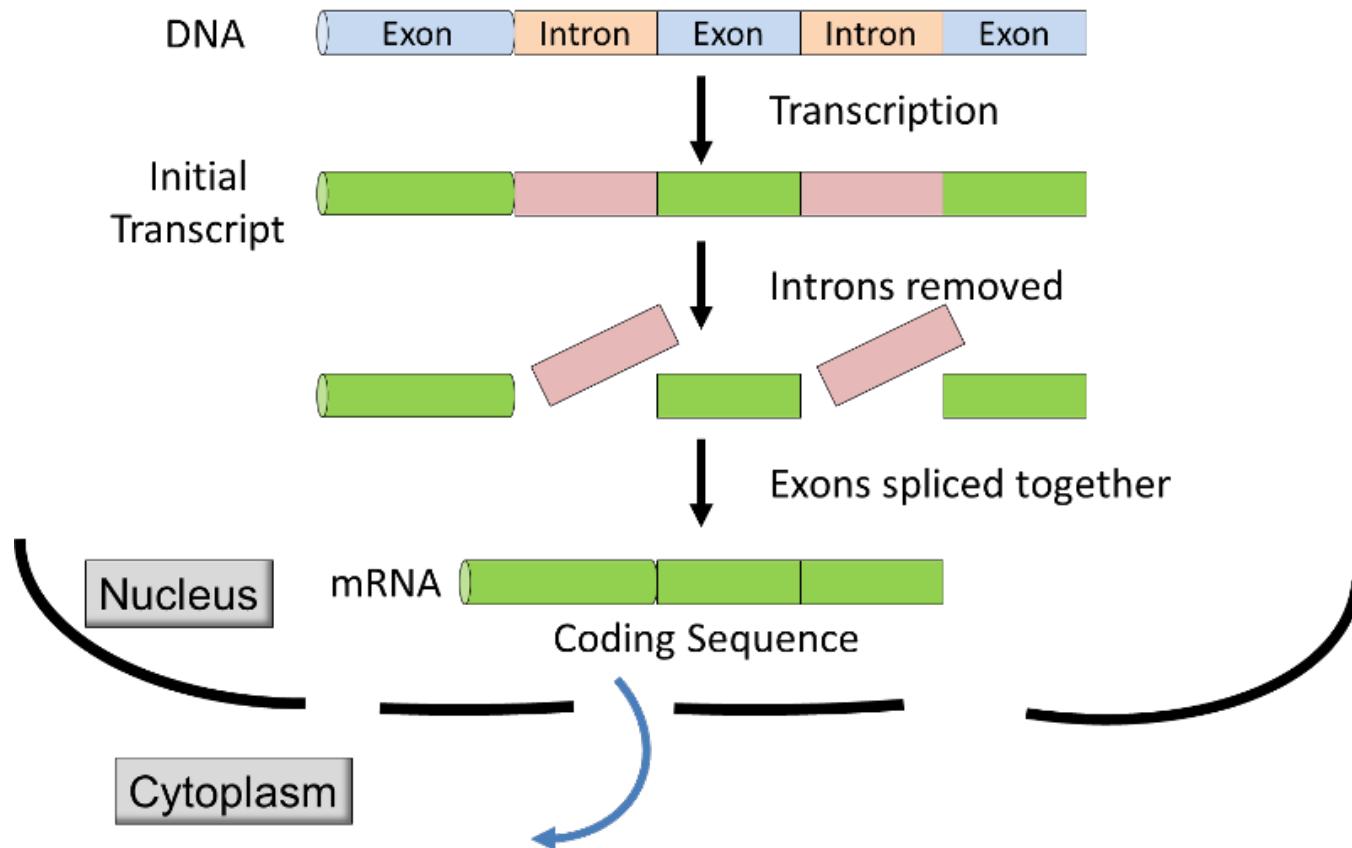
AACGCTGAGATCGCGAAATGGCGCTCGAGAGGCATAAGCGATCGAGCGAGCATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATA
TGCAGCGAGCTATCGATCGAGCGAGGCATTATTTCGAGGCAGGGCGAGCTATTATCGAGCGATGCGTAGCTAGTGATCGCTAGCTAGATGTCA
TTATATCGGGCGCGCGCAGCGAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTGATCGCTAGCTAGCTAGCGATATTATTAAATA
CTAGCATTATATAGTCGGCTAGTCAGTCGGCGCGCGATATTACGTAGCTAGCTAGCGATCGCTAGCTAGCGATCGTAGCTAGCTAG
TCAGTACGCTAGTCGATGCTAGCTAGCGATGCTAGCTAGCTAGCGATGCTATGTGATGCGTAGCTATTAGTCGATCGCGATGCTAG
CGATGCGATGCGTAGTCGATGCTAGCTAGCTAGCGATGCTATGTGATGCGTAGCTATTAGTCGATCGCGATGCTAG
CTAGTCGATGCTAGCTAGCATGCTAGATGTTAGTCGCGGATTATGCTAGCTAGTCGATGCTACGTAGCTTAGTCGACGTAG
TAGCTGATCGTAGCTAGCTAGTCGATGCTAGCTATGCGTAGCTATGCGTAGCGTATGGCTAACGCTGAGATCGCGAAATGGCGCT
CGAGAGGCTACGATCCAGGAGCATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATAATGCGAGCGAGCTATCGATCGAGCGA
GGCGATTATTCGAGGCCAGGCCTAGCTAGTCGAGCGAGGCAGCGAGCGATATTATTATATATCGGGCGCGCGCAGCG
CAGAGACGCGAAATATCGATCGATGCTAGCTAGCGATGCTAGCTAGTCGAGCTAGCTAGTCGAGCTAGCGATGCGTAGTCGATGCTAG
GTCGGCGCGCGCGCGCGATAATAGCTAGCTAGCGATGCTAGCTAGTCGAGCTAGCTAGTCGATGCGTAGTCGATGCTAG
ATCGATGCGATGCTAGCTAGCTAGCGATGCTAGCTAGTCGAGCTAGCTAGTCGAGCTAGTCGATGCGTAGTCGATGCTAG
CTAGCTAGCTATCGATCGATGCTAGTCGATGCTAGCTAGTCGAGCTAGCTAGTCGATGCGTAGTCGATGCTAGCTAGCGAT
CTAGATGTTATCTAGCTCGCGATTATGCTAGCTAGTCGATGCTACGCTAGCTAGTCGATGCTATGCTATCGAGCTA
GTCAGTCGCTAGCTATCGATCGCTATGCTATGCTATGGCTAACGCTGAGATCGCGAAATGGCGTAGCTAACGCTAAGCGATCGAGCGAG
CATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATAATGCGAGCGAGCTATCGATCGAGCTAACGCTAAGCGATCGAGCGAG
AGCTATTATCGCGAGCGAGGCAGGCAGCGAGCGATATTATTATATATCGGGCGCGCGCAGCGAGAGCGAGCGATG
CATCGTAGCTAGTGATCGCTAGCTAGATGTCAGTCAGCTAGCTAGTCGAGCTAGCTAGTCGATGCTAGTCGATGCTAGTCGATGCTAG
TACGTAGCTAGTCGAGCTAGCTAGCGATCGTAGCTAGTCGATGCTAGCTATCGATCGATGCGTAGTCGATGCTAGCTAGTCGATGCTAG
AGCTAGTCGATCGATGCTAGCGATCTAGCGATGCTAGCTAGTCGATGCGTAGTCGATGCTAGCTAGTCGATGCTAGTCGATGCTAG
TATGTGCATCGTAGCTATTAGTCGATCGCGATGCTAGCTAGTCGATGCTAGCTAGTCGATGCTAGCTAGTCGATGCTAGTCGATGCTAG
TGTCTAGCTAGTCGATGCTAGCTAGTCGACGCTAGCTAGTCGATGCTATGCTATCGAGCTAGTCGATGCTAGCTATGCTAGTCGATGCTAG
GCTATCGCTATGGCTAACGCTGAGATCGCGAAATGGCGCTCGAGAGGCTAACGCGATCGAGCGAGCGATCGAGCAGAGCGAGGGCGATCT
ATCGGCTATATTAAATAATGCGAGCGAGCTATCGATCGAGCGAGGCAGGGCGATTATTATCGAGGCAGGGCGAGCTATTATCGAGCGAGGC
GCGAGCGATATTATTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTCGATGCTAG
CTAGATGTCAGTCAGCTAGCATTATAGTCGCGCTAGTCAGTCGCGCGCGCGAGCTATTACGTAGCTAGTCGAGCTAGCTAG
CGATCGTAGCTAGCTAGTCAGTCAGCTAGTCGATGCTAGCTAGTCGATGCTAGCTAGTCGATGCTAGTCGATGCTAG
TCTAGCGATGAGCTATCGATCGCGATGCGTAGTCGATGCTAGCTAGCTAGTCGATGCTAGTCGATGCTAGTCGATGCTAG
GATCGCGCGATGCTAGCTAGTCGATGCTAGCTAGCATGCTAGATGTTATCTAGCTCGCGGATTATGCTAGCTAGTCGATGCTACGTA
GCTTAGTCGACGCTAGCTAGTCGATGCTAGCTAGTCGATGCTAGTCGATGCTAGTCGATGCTAGTCGATGCTAGTCGATGCTAG

X 2.000.000 times!

DNA contains “functional” information



DNA contains “functional” information



DNA contains “functional” information

gene (functional unit)

```
...AACGCTGAGATCGCGAAATGGCGCTCGAGAGGGCTA  
AGCGATCGAGCGAGCATTGAGCAGAGCGAGGGCGAT  
CTATCGGCTATTAATATGCGAGCGAGCTATCGATCG  
AGCGAGGCAGATTATTATTCAAGGCGAGGCAGCTATT  
CATGGCGAGCGAGGCAGCGAGCGATATTATTATTAT  
ATATCGGGCGCGCGCGCAGCGCAGAGACCGAATTA  
TCGAGCGATGCATGCGTAGCTAGTGTATCGGCTAGCTA  
GATGTCAGTGCAGCTAGCATTATATAGTGCGGCTAGTC  
AGTCGCGGGCGCGCGCGATATACGTAGCTAG...
```

A: brown eyes



```
...AACGCTGAGATCGCGAAATGGCGCTCGAGAGGGCTA  
AGCGATCGAGCGAGCATTGAGCAGAGCGAGGGCGAT  
CTATCGGCTATTAATATGCGAGCGAGCTATCGATCG  
AGCGAGGCAGATTATTATTCAGAGGCGAGGCAGCTATT  
CATGGCGAGCGAGGCAGCGAGCGATATTATTATTAT  
ATATCGGGCGCGCGCGCAGCGCAGAGACCGAATTA  
TCGAGCGATGCATGCGTAGCTAGTGTATCGGCTAGCTA  
GATGTCAGTGCAGCTAGCATTATATAGTGCGGCTAGTC  
AGTCGCGGGCGCGCGCGATATACGTAGCTAG...
```

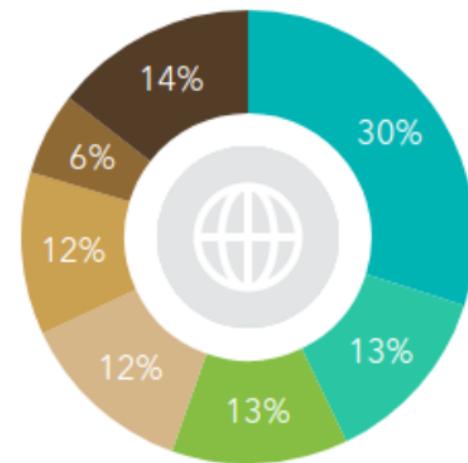
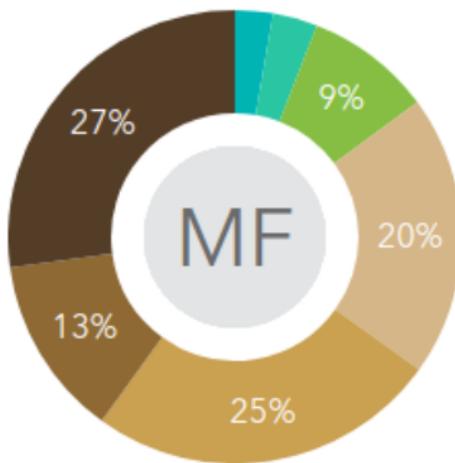
G: blue eyes



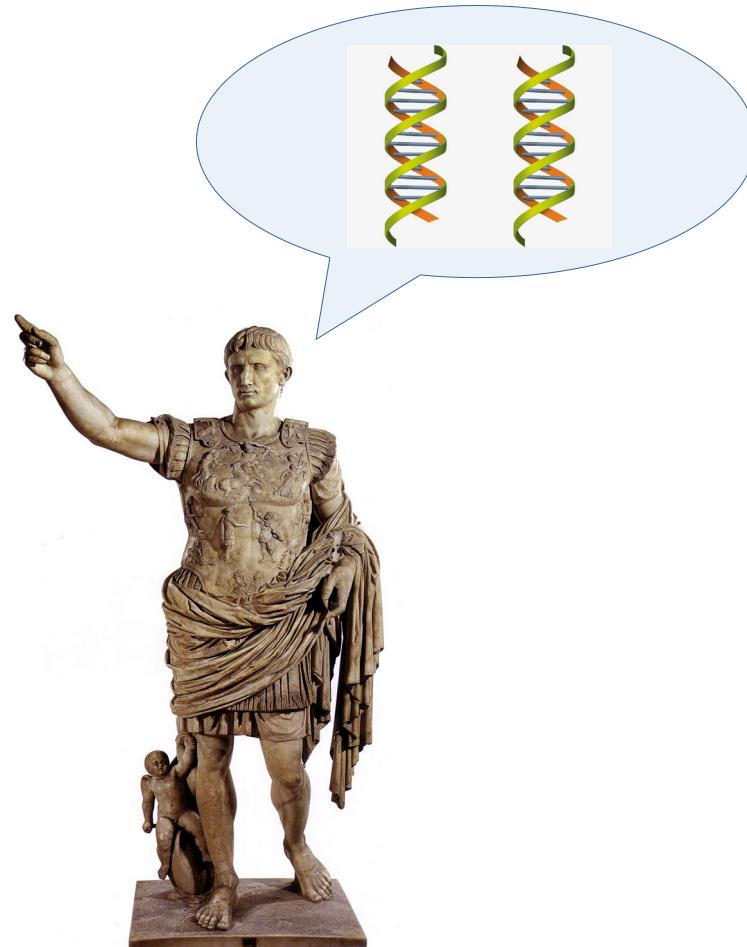
DNA contains “functional” information

Matteo, you are likely to have dark-colored eyes.

65% of customers who are genetically similar to you have dark hazel, light brown, or dark brown eyes.



DNA contains “historical” information



How much does DNA sequencing cost?

1990-2000: Human Genome project:
10M USD

2016: 100,000 Genomes Project UK: 1,000 USD

2012-: personal DNA kits: 100 USD

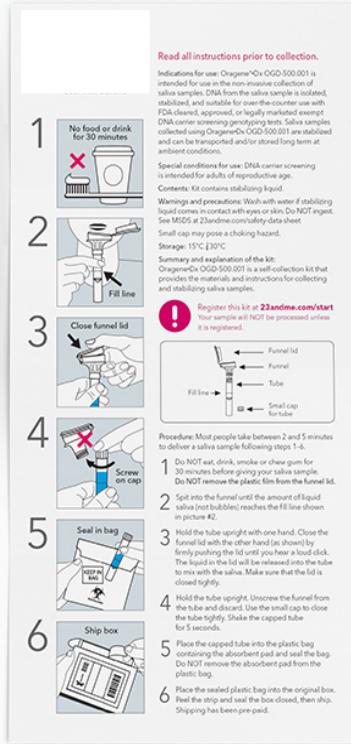
My DNA



saliva collection kit



specimen bag



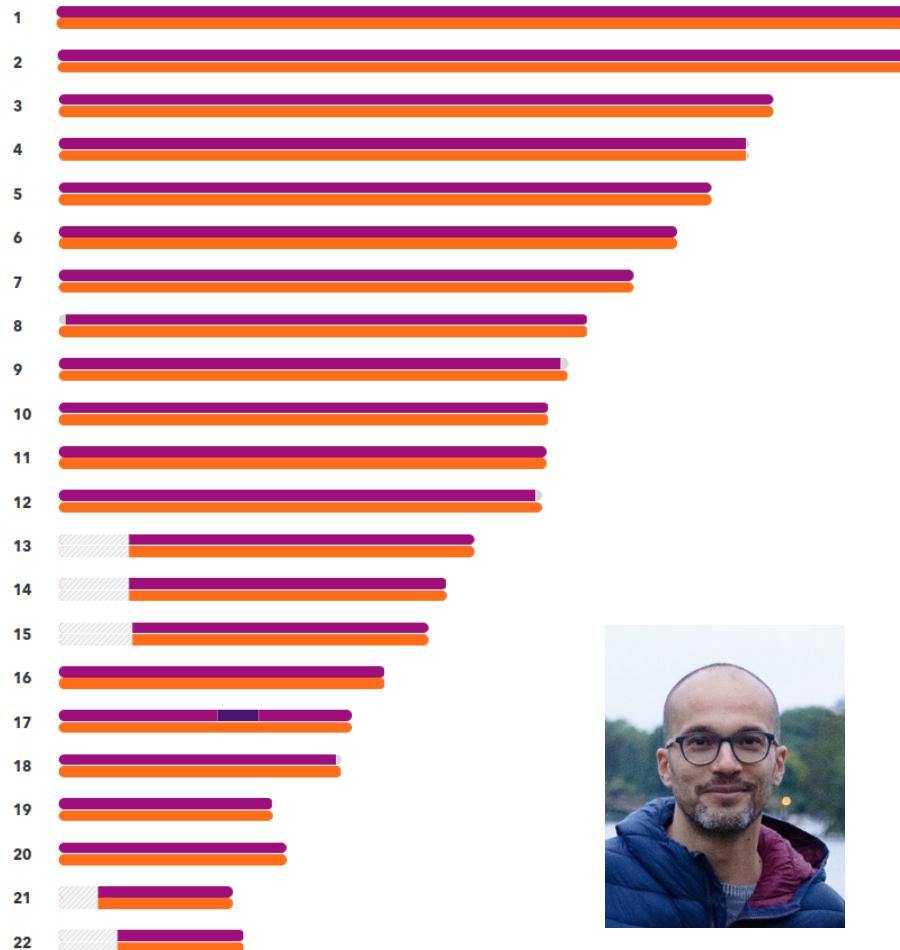
step by step instructions



tube container

PS:I am not endorsing at all the business model behind this or similar companies

My DNA



DNA contains “historical” information

Pair of chromosomes



DNA contains “historical” information

dad

mum

Pair of chromosomes



DNA contains “historical” information



+



=



DNA contains “historical” information

25%



+ Paternal
Grandpa



+ Paternal
Grandma

25%



+ Maternal
Grandpa

25%



+ Maternal
Grandma

25%



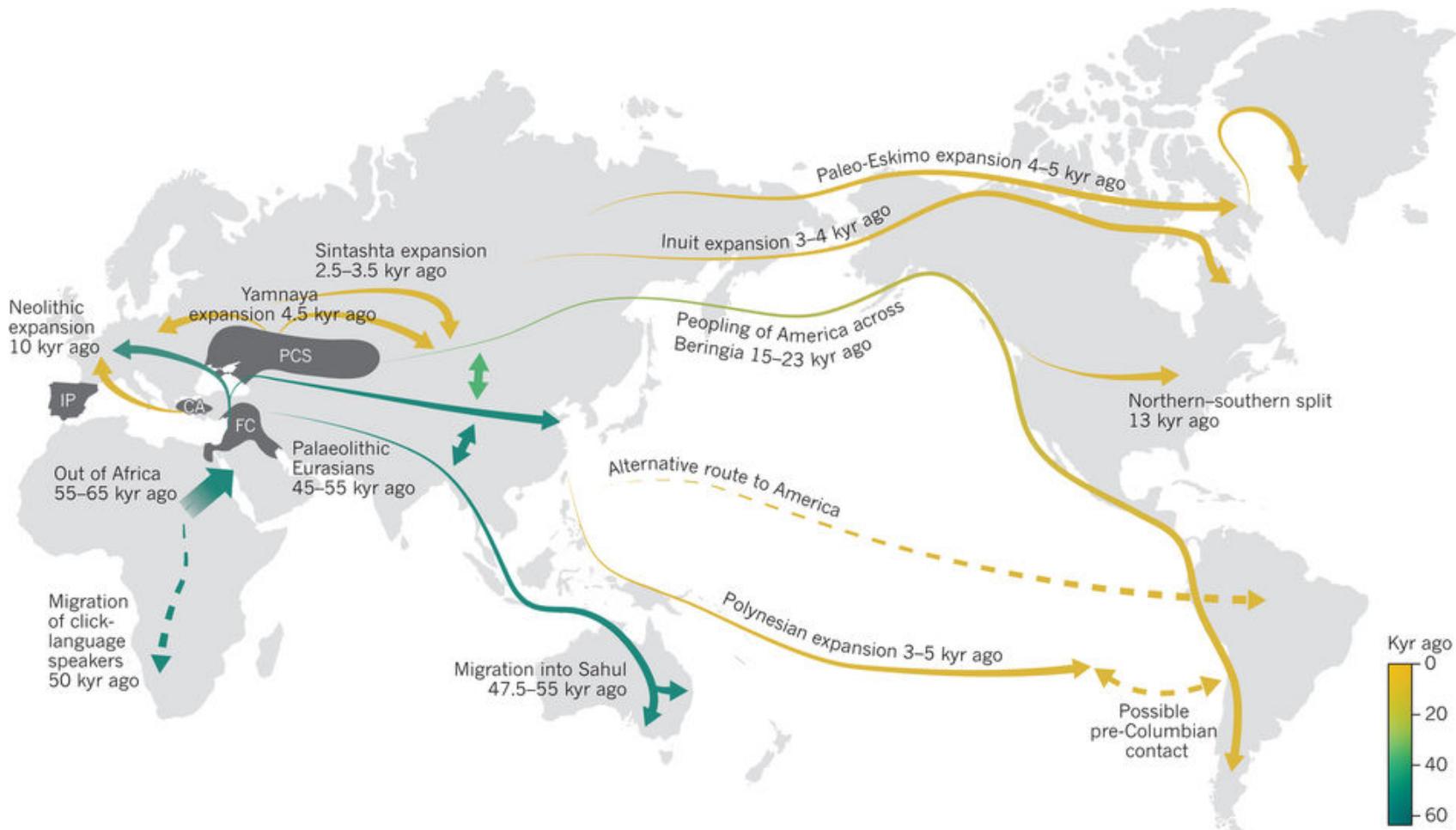
~50% DNA from
Adelio Fumagalli



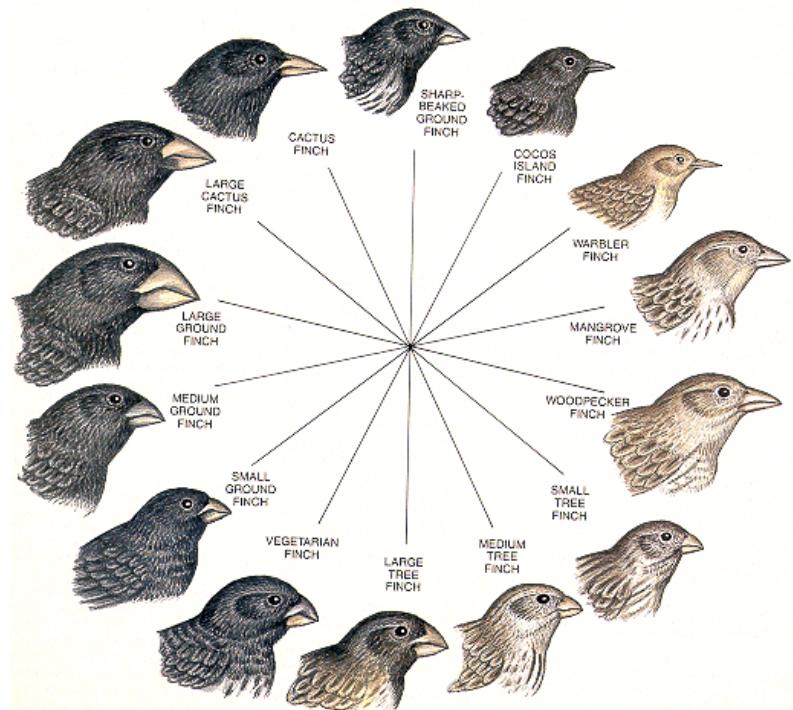
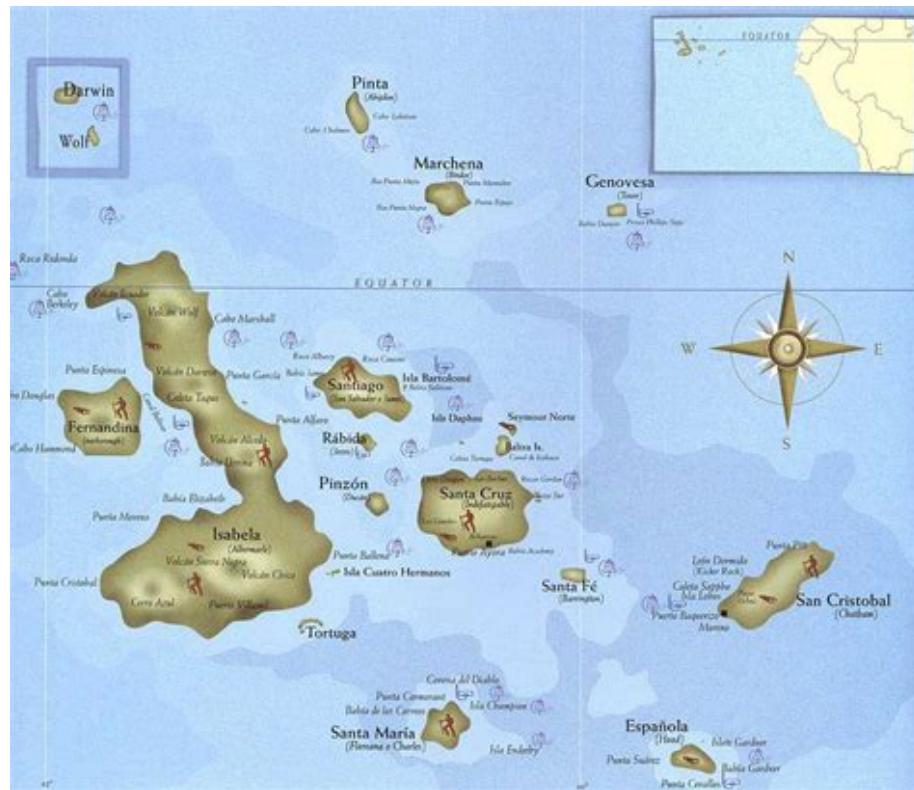
Matteo Fumagalli

~50% DNA from
Maria Assunta
Cattaneo

Demographic inferences: human evolution



Inferences of natural selection: how species adapted to new/changing environments



Applications of genomics in evo/eco/cons!



Genomics and Bioinformatics

How about Bioinformatics?

What is Bioinformatics?

What is bioinformatics?

Bioinformatics /baɪ.oʊ,ɪnfər'mætɪks/ (listen) is an interdisciplinary field that develops methods and software tools for understanding biological data, in particular when the data sets are large and complex. As an interdisciplinary field of science, bioinformatics combines biology, computer science, information engineering, mathematics and statistics to analyze and interpret the biological data. Bioinformatics has been used for *in silico* analyses of biological queries using mathematical and statistical techniques. [clarification needed]

from wikipedia

“From Data to Knowledge” (Professor Janet Thornton)

Genomics and bioinformatics

Intended Learning Outcomes:

- understand the theory of population genomics and its applications to estimate evolutionary parameters
- describe how genomic data is generated and the challenges associated to its analysis
- implement scripts in R to perform evolutionary inferences from genomic data

Genomics and Bioinformatics

A useful primer in genomics can be found here:

<http://compgenomr.github.io/book/intro.html>

I advice students who didn't study biology in their undergraduate course to go over chapters 1.1 and 1.3 as a gentle introduction to some basic concepts in genomics we will discuss during this module.

This information is also present on blackboard.

In videos, please ignore any reference to a bitbucket website. All data is available on blackboard.

Use the Teams chat to ask questions during either the video lecture, Q&A or practical (vote if you wish to ask the same question).