

# Genomics and bioinformatics

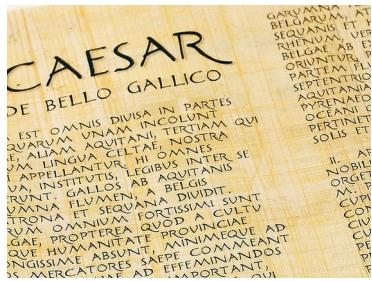
Introduction on evo/eco/cons inferences from genomic data using R

Matteo Fumagalli

# Using genetics to infer the evolutionary history of a species

**History + Genetics = ???**

# History + Genetics = ???



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

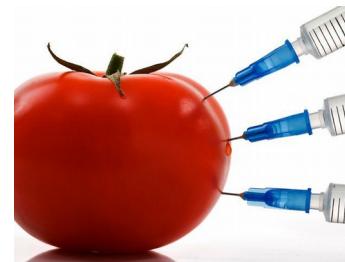
[Reply](#) [Retweet](#) [Favorite](#) [More](#)

1,105 RETWEETS 863 FAVORITES



6.37 PM - 8 May 13

# History + Genetics = ???



Donald J. Trump  [@realDonaldTrump](#)

[Follow](#)

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

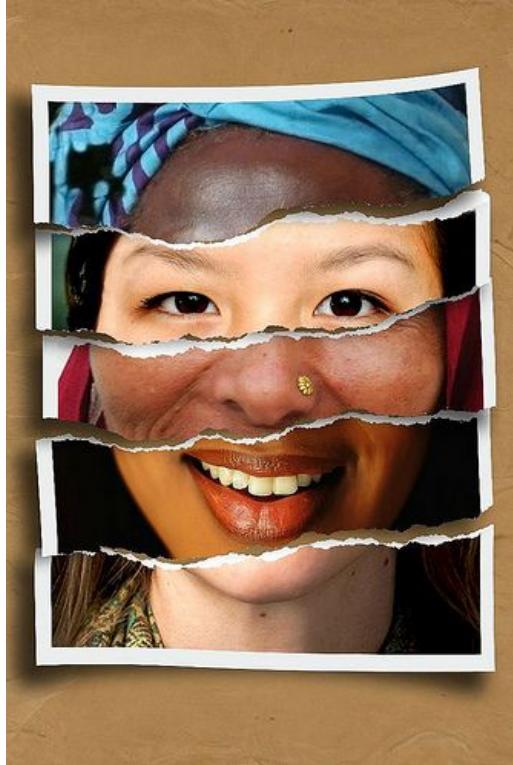


6:37 PM - 8 May 13

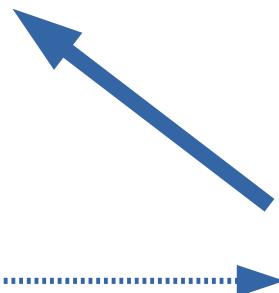


How can we use genetics to learn something about the evolutionary history of a species (or population)?

# Everyone of us is unique



You



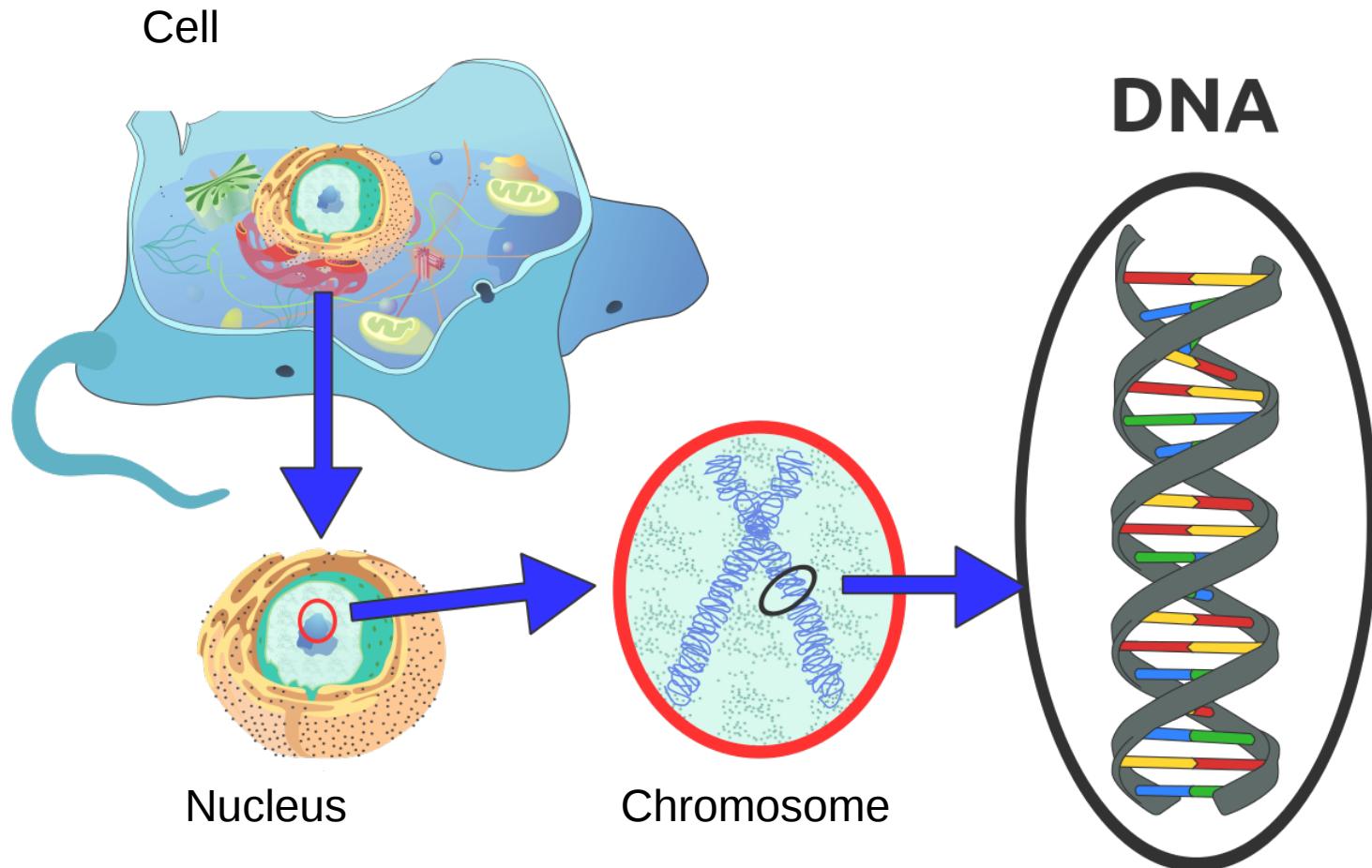
DNA



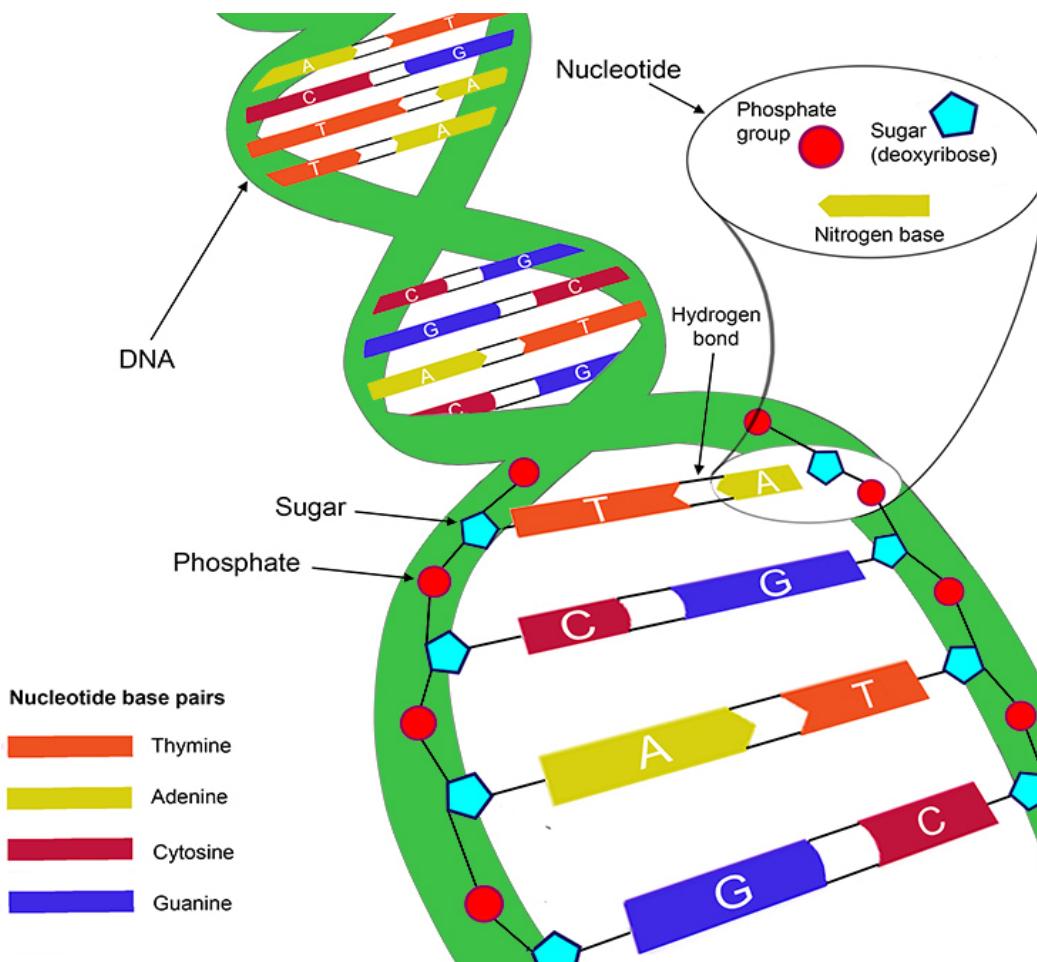
Environment



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)

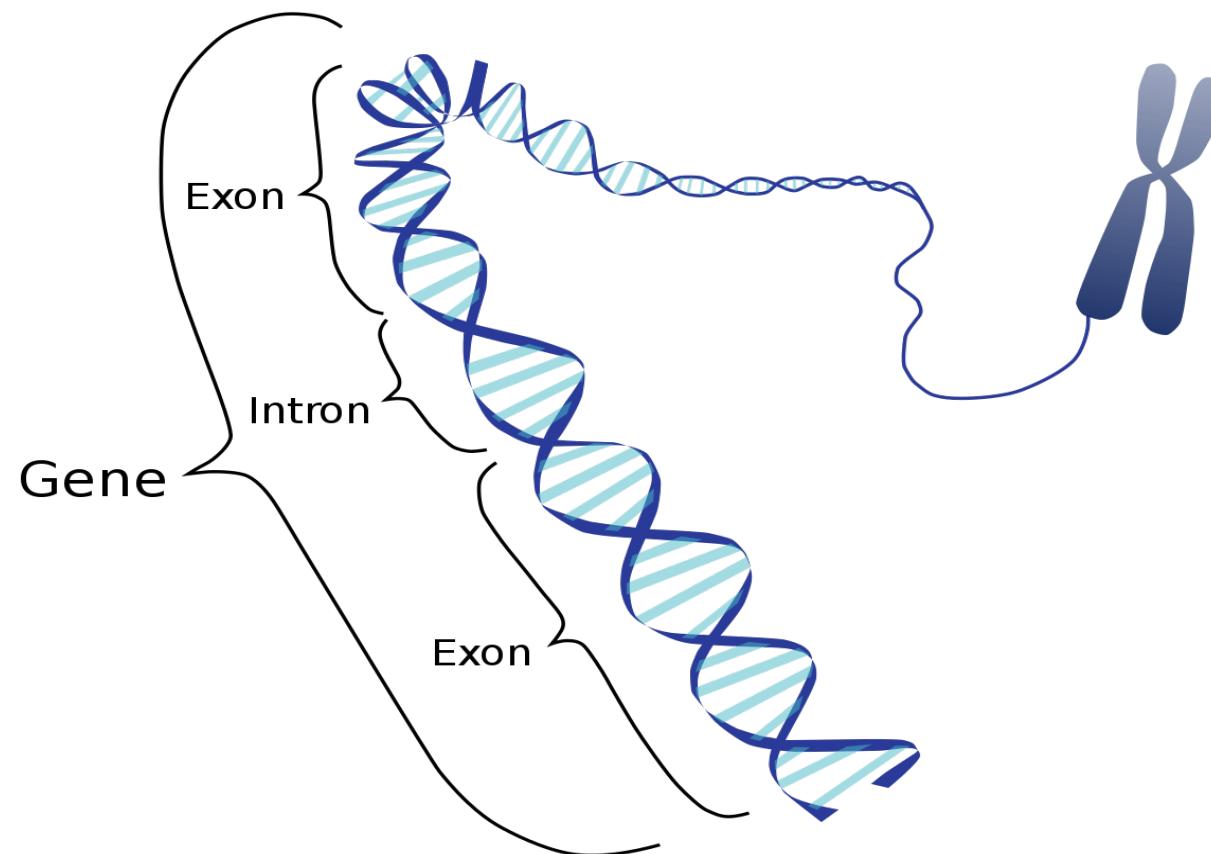
AACGCTGAGATCGCGAAATGGCGCTCGAGAGGCATAAGCGATCGAGCGAGCATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATA  
TGCAGCGAGCTATCGATCGAGCGAGGCATTATTTCGAGGCAGGGCGAGCTATTATGGCGAGCGAGGCAGCGATATTATTAA  
TTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTGATCGCTAGCTAGATGTCACTGAG  
CTAGCATTATATACTGCGGCTAGTCAGTCGGCGCGCGATATTACGTAGCTAGCTAGCGATCGTAGCTAGCTAGCGATGTAGCTAG  
TCAGTACGCTAGTCGATGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGATGTAGCTAT  
CGATGCGATGCGTAGTCGATGCTAGCTAGCTAGCGATGCGTAGCTATGCTATGCGATGCGTAGCTATTAGTCGATCGCGATGCGTAG  
CTAGTCGATGCTAGCTAGCATGCTAGATGTTATCTAGCTCGGGATTATGCTAGCTAGTCGATGCTACGTAGCTTAGTCGACGTAGC  
TAGCTGATCGCTATGCTATCGAGCTAGTCGCTAGCTATGCGATGCGTAGCTATGCGTAGCTGGCTAACGCTGAGATCGCGAAATGGCGCT  
CGAGAGGCTAACGCGATCGAGCGAGCATCGAGCAGAGCGAGGGCGATCTATGGCTATATTAAATATGCGAGCGAGCTATCGATCGAGCGA  
GGCGATTATTATTGAGGCAGGGCGAGCTATTATGGCGAGCGAGGCAGCGAGCGATATTATTATTATATCGGGCGCGCGCAGCG  
CAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTGATCGCTAGCTAGCGATGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAG  
GTCGGCGCGCGCGCGATATTACGTAGCTAGCGATGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAG  
ATCGATGCGATGCTAGCTAGCTAGCGATGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGTAG  
CTAGCTAGCTATCGATCGATGCGATGCTAGCTAGCGATGCGTAGCTATTAGTCGATCGCGATGCGTAGCTAGCGATGCGTAG  
CTAGATGTTATCTAGCTCGGGATTATGCTAGCTAGTCGATGCTACGTAGCTAGCTAGCTGCTATGCTATCGAGCTA  
GTCAGTCGCTAGCTATGCGATGCGTAGCTATGCGTAGCTGGCTAACGCTGAGATCGCGAAATGGCGCTCGAGAGGCTAACGCGATCGAGCGAG  
CATCGAGCAGAGCGAGGGCGATCTATGGCTATATTAAATATGCGAGCGAGCTATCGATCGAGCGAGGGCGATTATTATTGAGGCAGGGCG  
AGCTATTATGGCGAGCGAGGCAGGCAGCGAGCGATATTATTATTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATG  
CATCGTAGCTAGTGATCGGCTAGCTAGATGTCAGTCAGCTAGCATTATATAGTCGCGCTAGTCAGTCGCGCAGCGTAGCTAGCGATG  
TACGTAGCTAGCGAGCTAGCGATCGTAGCTAGCTAGCGATGCGTAGCTAGCTATCGATGCGATGCGTAGCTAGCTAGCGATG  
AGCTAGTCGATCGATGCTAGCGATCTAGCGATGCGTAGCTAGCTATCGATGCGATGCGTAGCTAGCTAGCTAGCTATCGATGCGATG  
TATGTCATGCGTAGCTATTAGTCGATCGCGCGATGCGTAGCTAGTCGATGCTAGCTAGCTAGCATGCTAGATGTTATCTAGCTCGGGATT  
TGTCTAGCTAGTCGATGCTACGTTAGTCGACGCTAGCTAGCTGATCGCTATGCTATCGAGCTAGTCGCTAGCTATGCGATGCGTAT  
GCTATGCGTAGGGCTAACGCTGAGATCGCGAAATGGCGCTCGAGAGGCTAACGCGATCGAGCGAGCGATCGAGCAGAGCGAGGGCGATCT  
ATCGGCTATATTAAATATGCGAGCGAGCTATCGATCGAGCGAGGCAGGGCGATTATTATTGAGGCAGGGCGAGCTATTATGGCGAGCGAGGC  
GCGAGCGATATTATTATTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTGATCGGCTAG  
CTAGATGTCAGTCAGCTAGCATTATATAGTCGCGCTAGTCAGTCGCGCGCGCGAGCTATTACGTAGCTAGTCGAGCTAGCTAG  
CGATCGTAGCTAGCTAGCGTAGTCGATGCTAGCTGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCTAGTCGATGCGTAGCTAGCGA  
TCTAGCGATGAGCTATCGATGCGATGCGTAGTCGATGCTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTATTAGTC  
GATCGCGCGATGCTAGCTAGTCGATGCTAGCTAGCATGCTAGATGTTATCTAGCTCGCGGATTATGCTAGCTAGTCGATGCTACGTA  
GCTTAGTCGACGCTAGCTAGCTGATCGCTATCGAGCTAGTCGATGCTAGCTAGCGATGCGTAGCTATGCGTAGCTAGCGTAGCTAGCGA

## DNA (genome)

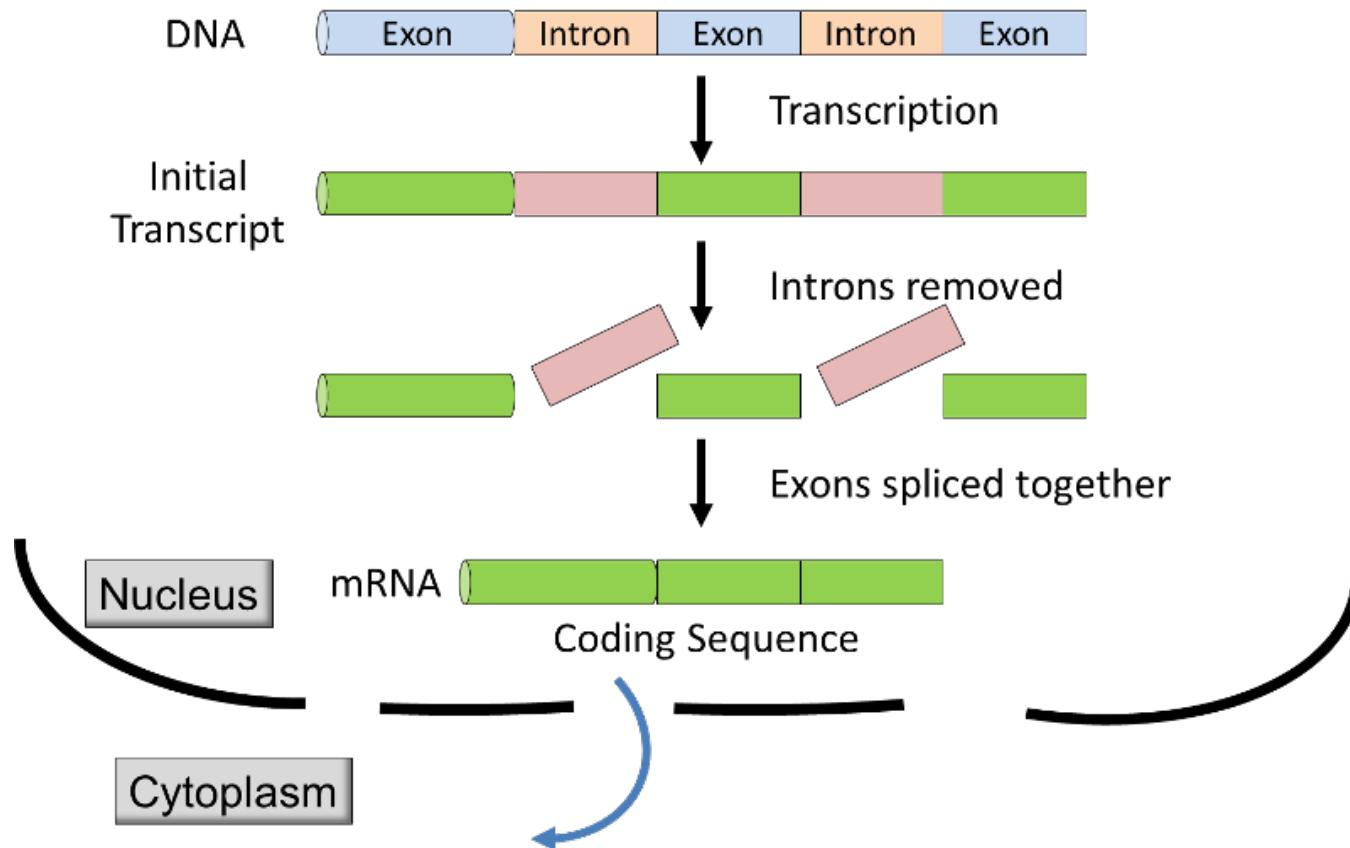
AACGCTGAGATCGCGAAATGGCGCTCGAGAGGCATAAGCGATCGAGCGAGCATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATA  
TGCAGCGAGCTATCGATCGAGCGAGGCATTATTTCGAGGCAGGGCGAGCTATTATCGAGCGATGCGTAGCTAGTGATCGCTAGCTAGATGTCA  
TTATATCGGGCGCGCGCAGCGAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTGATCGCTAGCTAGCTAGCGATATTATTAAATA  
CTAGCATTATATAGTCGGCTAGTCAGTCGGCGCGCGATATTACGTAGCTAGCTAGCGATCGCTAGCTAGCGATCGTAGCTAGCTAG  
TCAGTACGCTAGTCGATGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGATGCGATGCGTAGCTAT  
CGATGCGATGCGTAGTCGATGCTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAG  
CTAGTCGATGCTAGCTAGCATGCTAGATGTTATCTAGCTCGGGATTATGCTAGCTAGTCGATGCTACGTAGCTAGTCGACGTAGC  
TAGCTGATCGTAGCTAGCTAGCGATGCTAGCTAGCGTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGAAATGGCGCT  
CGAGAGGCTAATCGATCCAGGAGCATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATAATGCGAGCGAGCTATCGATCGAGCGA  
GGCGATTATTGAGGCCAGGCCTGCTATCGGCGAGCGAGGCAGCGAGCGATATTATTATATATCGGGCGCGCGCAGCG  
CAGAGACGCGAAATATCGATCGATGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCG  
GTCGGCGCGCGCGCGCGATAATAGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCG  
ATCGATGCGATGCTAGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAG  
CTAGCTAGCTATCGATCGATGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCG  
CTAGATGTTATCTAGCTCGGGATTATGCTAGCTAGTCGATGCTACGCTAGCTAGCTAGCTAGCGCTATGCTATCGAGCTA  
GTCAGTCGCTAGCTATCGATCGCTATGCTAGCTAGCGTAGCTAGCTAGCGATCGCTAGCTAGCGATCGAGCGAG  
CATCGAGCAGAGCGAGGGCGATCTATCGGCTATATTAAATAATGCGAGCGAGCTATCGATCGAGCGATCGAGCG  
AGCTATTATCGCGAGCGAGGCAGGCAGCGAGCGATATTATTATATATCGGGCGCGCGCAGCGAGAGCGAGCG  
CATCGTAGCTAGTGATCGCTAGCTAGCTAGTGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCGATGCG  
TACGTAGCTAGCGAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAG  
AGCTAGTCGATCGATGCTAGCGATCTAGCGATGCTAGCTATCGATCGATGCGTAGCTAGCGATGCGTAGCTAGCGATGCG  
TATGTGCATCGTAGCTATTAGTCGATCGCGATGCTAGCTAGTCGATGCTAGCTAGCGATGCGTAGCTAGCGATGCG  
TGTCTAGCTAGTCGATGCTACGCTAGCTAGTCGACGCTAGCTAGCTAGTCGATGCTATGCTATCGAGCTAGTCG  
GCTATCGCTATGGCTAACGCTGAGATCGGAAATGGCGCTCGAGAGGCTAACGCGATCGAGCGAGCGATCGAGCG  
ATCGGCTATATTAAATAATGCGAGCGAGCTATCGATCGAGCGAGGCAGGGCGATTATTATCGAGGCAGGGCGAGCTATT  
CGAGCGATATTATTATATCGGGCGCGCGCAGCGCAGAGACGCGAATTATCGAGCGATGCGTAGCTAGTGATCG  
CTAGATGTCAGTCAGCTAGCATTATAGTCGGCTAGTCAGTCGCGCGCGCGATATTACGTAGCTAGTCGAGCTAGCTAG  
CGATCGTAGCTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCTAGCGATGCGTAGCTAGCG  
TCTAGCGATGAGCTATCGATCGCGATGCGTAGTCGATGCTAGCTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTATTAG  
GATCGCGCGATGCTAGCTAGTCGATGCTAGCTAGCGATGCTAGATGTTATCTAGCTCGGGATTATGCTAGCTAGTCGATGCTACGTA  
GCTTAGTCGACGCTAGCTAGCTAGCGCTATGCTAGCGTAGCTAGCGTAGCTAGCGATGCGTAGCTAGCGATGCGTAGCTAGCGA

**X 2.000.000 times!**

## DNA contains “biological” information



## DNA contains “biological” information



## DNA contains “functional” information

gene (functional unit)

```
...AACGCTGAGATCGCGAAATGGCGCTCGAGAGGGCTA  
AGCGATCGAGCGAGCATTGAGCAGAGCGAGGGCGAT  
CTATCGGCTATTAATATGCGAGCGAGCTATCGATCG  
AGCGAGGCAGATTATTATTCAAGGCAGGGCGAGCTATT  
CATGGCGAGCGAGGGCGAGCGAGCGATATTATTATTAT  
ATATCGGGCGCGCGCGCAGCGCAGAGACCGAATTA  
TCGAGCGATGCATGCGTAGCTAGTGTATCGGCTAGCTA  
GATGTCAGTGCAGCTAGCATTATATAGTGCGGCTAGTC  
AGTCGCGGGCGCGCGCGATATACGTAGCTAG...
```

A: brown eyes



```
...AACGCTGAGATCGCGAAATGGCGCTCGAGAGGGCTA  
AGCGATCGAGCGAGCATTGAGCAGAGCGAGGGCGAT  
CTATCGGCTATTAATATGCGAGCGAGCTATCGATCG  
AGCGAGGCAGATTATTATTCAGAGGCAGGGCGAGCTATT  
CATGGCGAGCGAGGGCGAGCGAGCGATATTATTATTAT  
ATATCGGGCGCGCGCGCAGCGCAGAGACCGAATTA  
TCGAGCGATGCATGCGTAGCTAGTGTATCGGCTAGCTA  
GATGTCAGTGCAGCTAGCATTATATAGTGCGGCTAGTC  
AGTCGCGGGCGCGCGCGATATACGTAGCTAG...
```

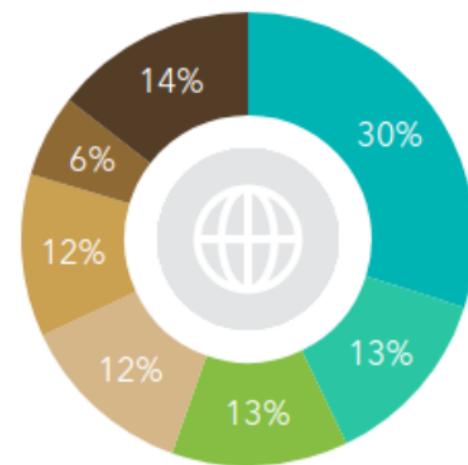
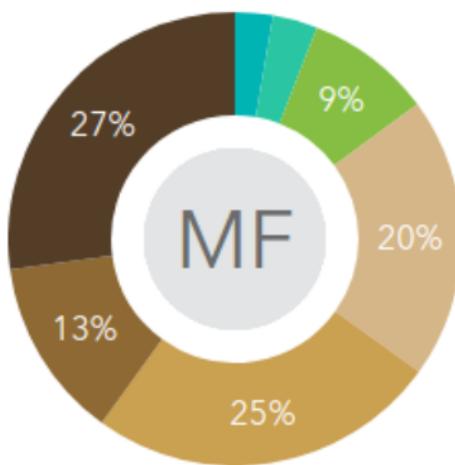
G: blue eyes



## DNA contains “functional” information

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

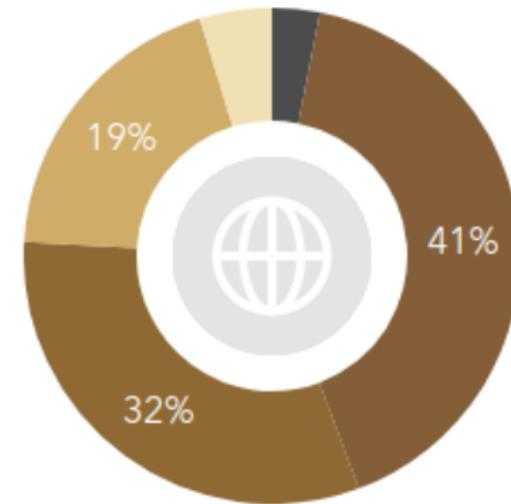
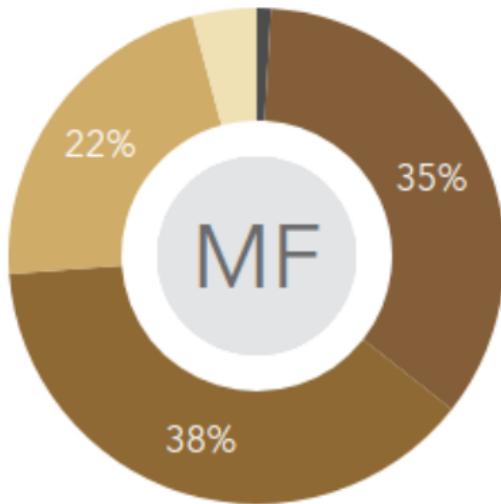
65% of customers who are genetically similar to you have dark hazel, light brown, or dark brown eyes.



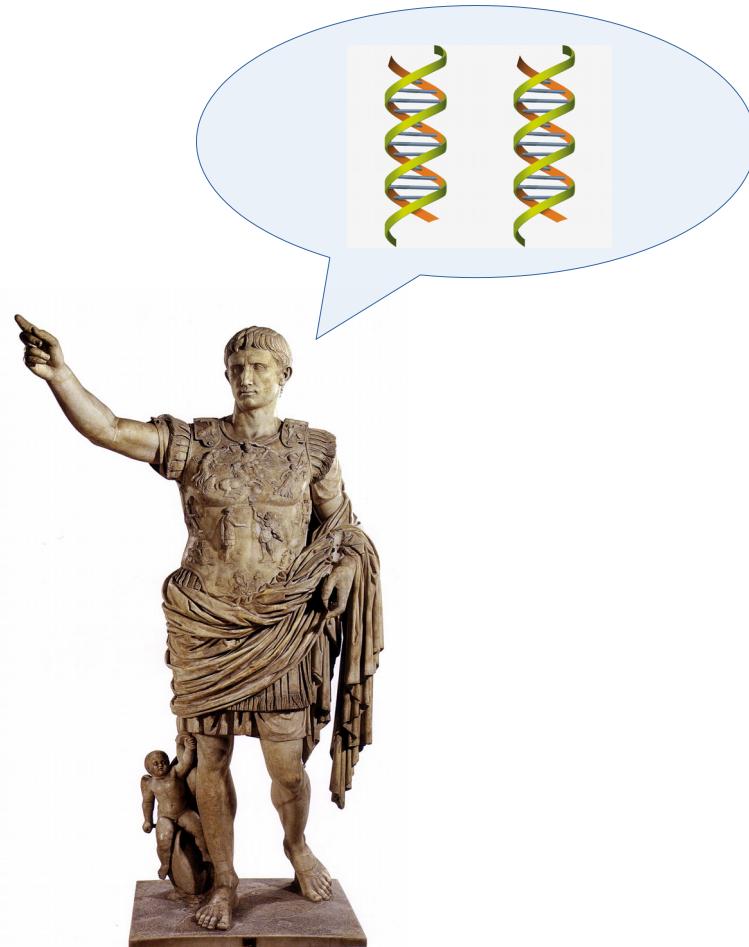
## DNA contains “functional” information

Matteo, you are likely to have light hair.

64% of customers who are genetically similar to you have light brown or blond hair.



## 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

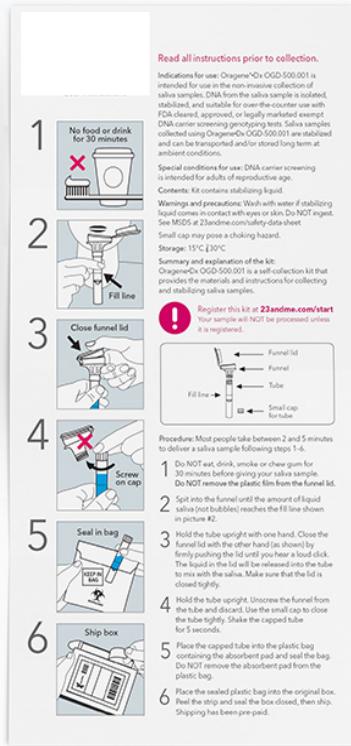
## My DNA



saliva collection kit



specimen bag

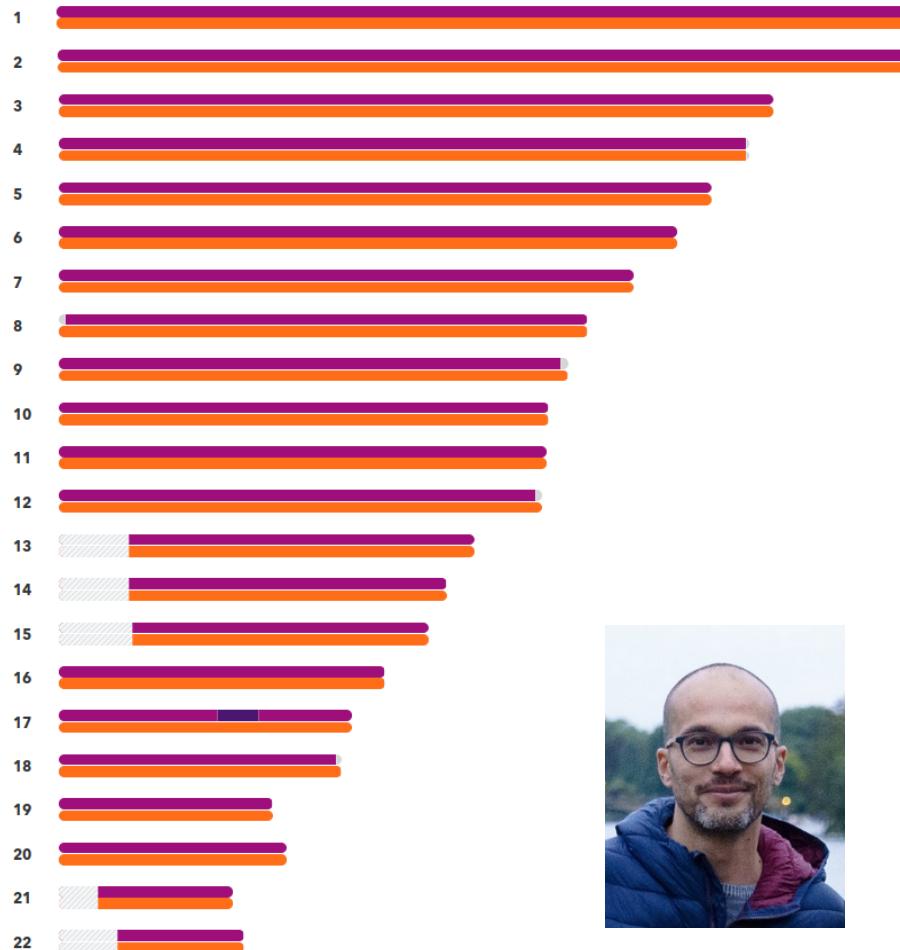


step by step instructions



tube container

# My DNA



# DNA contains “historical” information

Pair of chromosomes



# DNA contains “historical” information

dad

mum

Pair of chromosomes



# DNA contains “historical” information



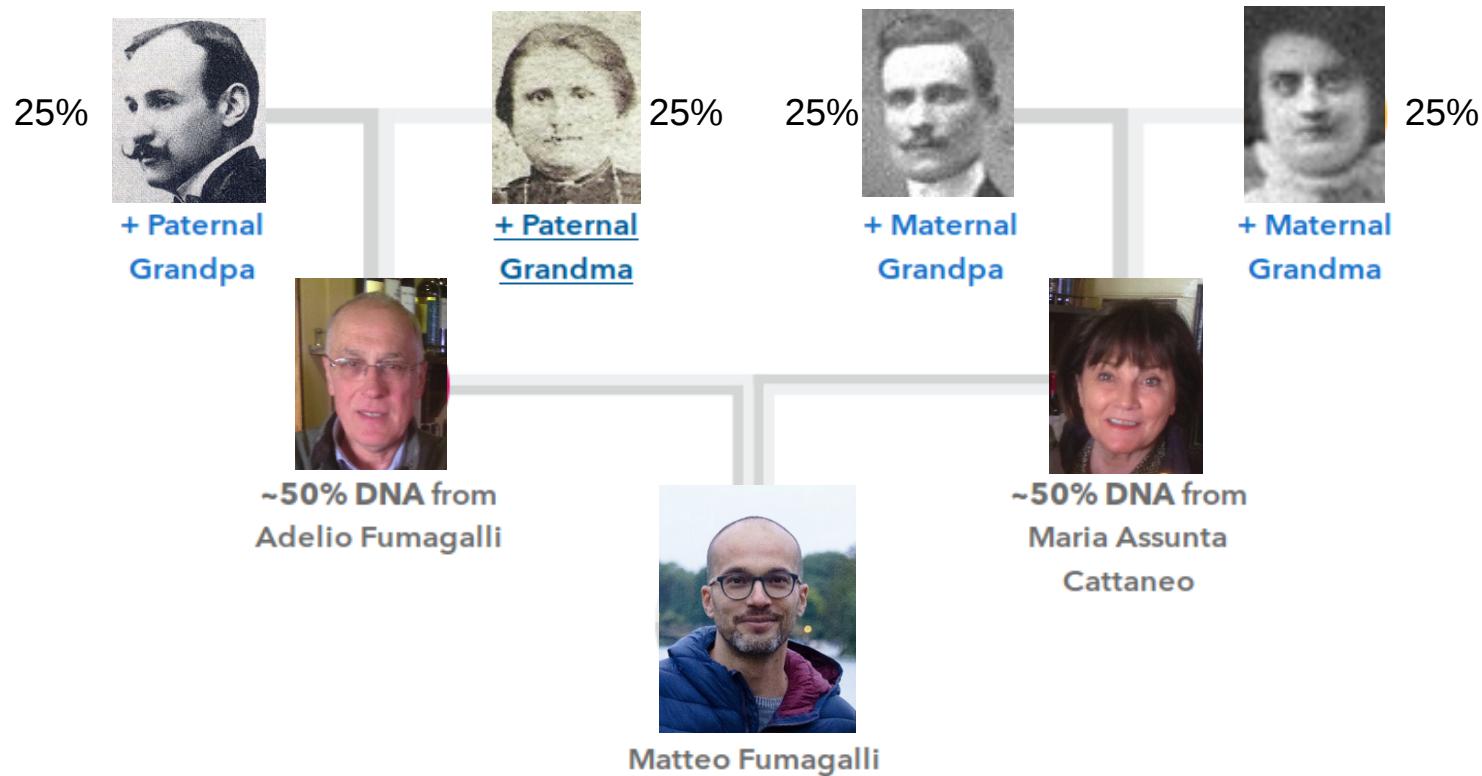
+



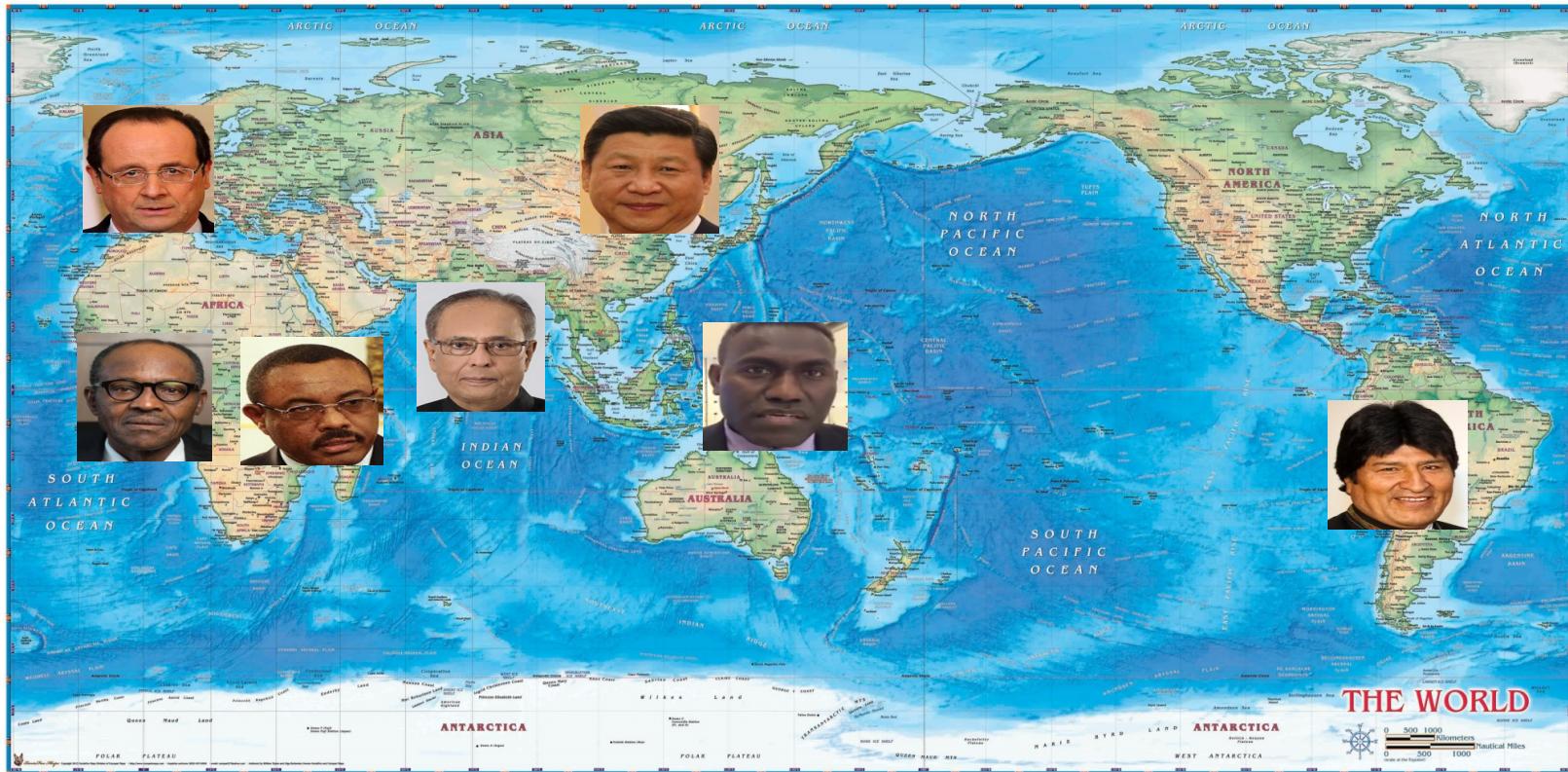
=



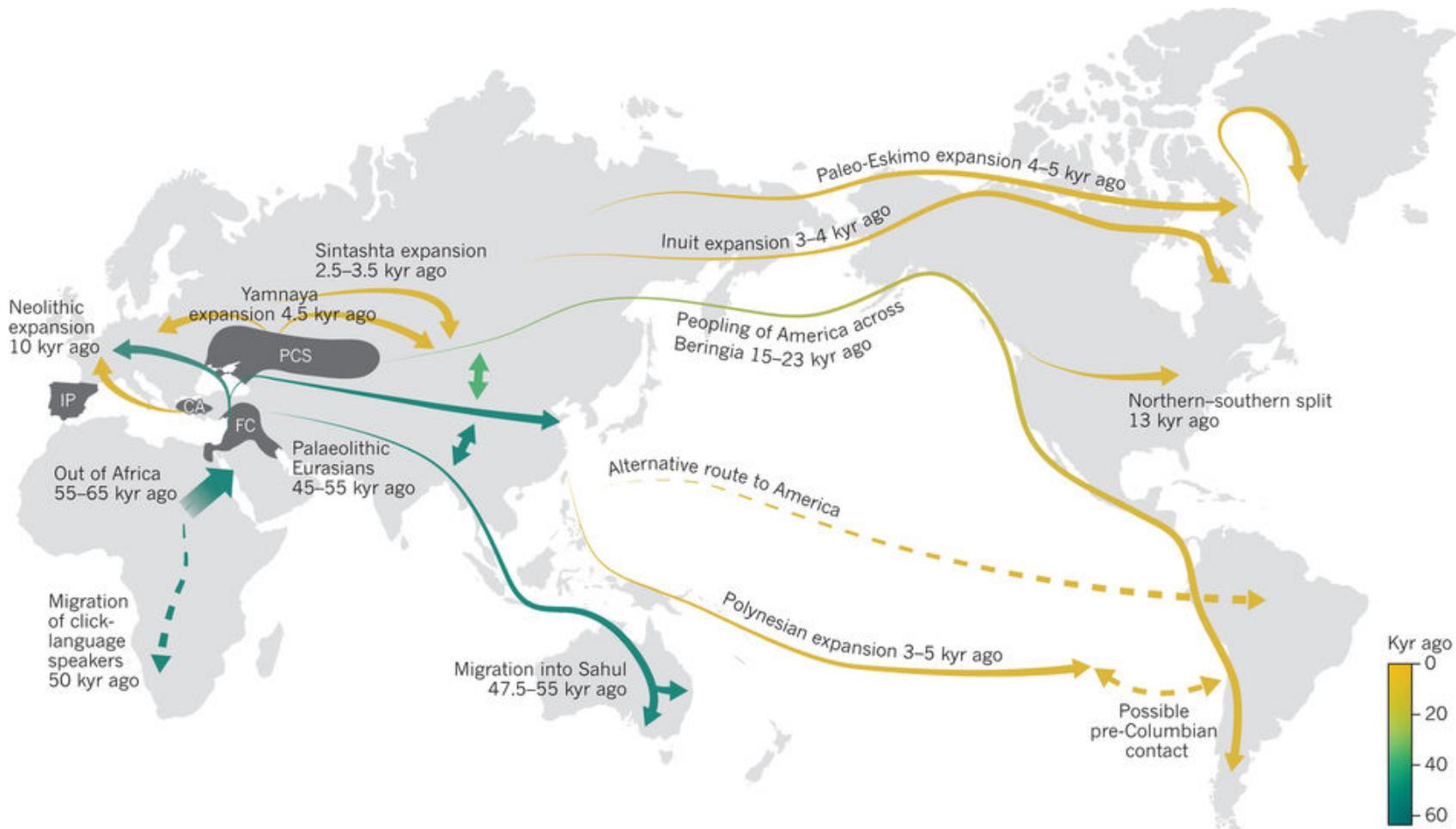
# DNA contains “historical” information



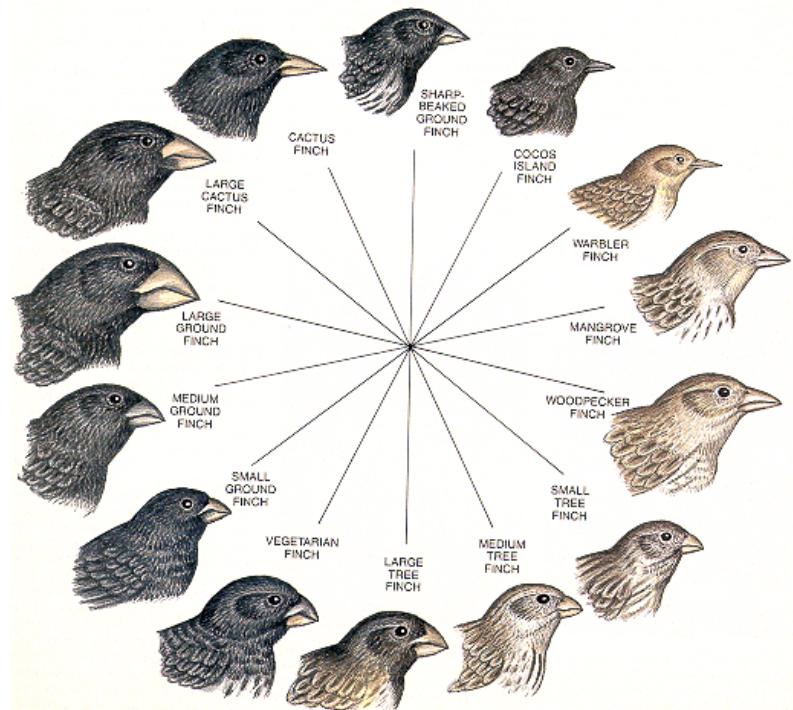
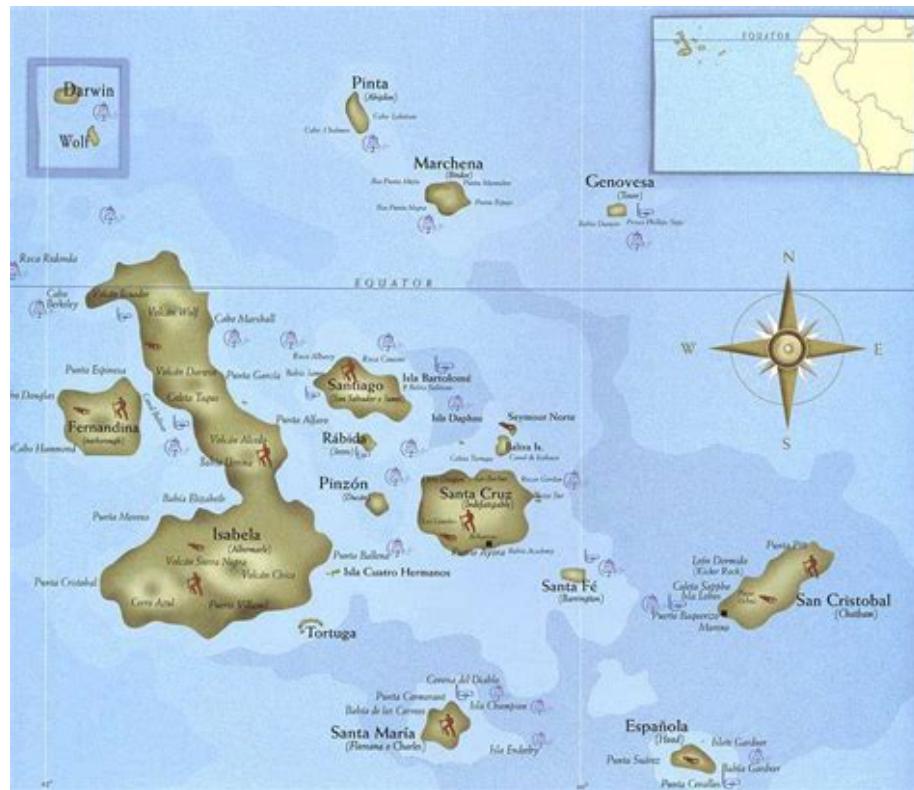
# Demographic inferences



# Demographic inferences: human evolution



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



## Applications



# Genomics and bioinformatics

## Intended Learning Outcomes:

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