

Biogeography of arabidopsis

students:

Anton Eliseev

Kristina Krivosova

the “pahan”

Tatiana Tatarinova

The goal

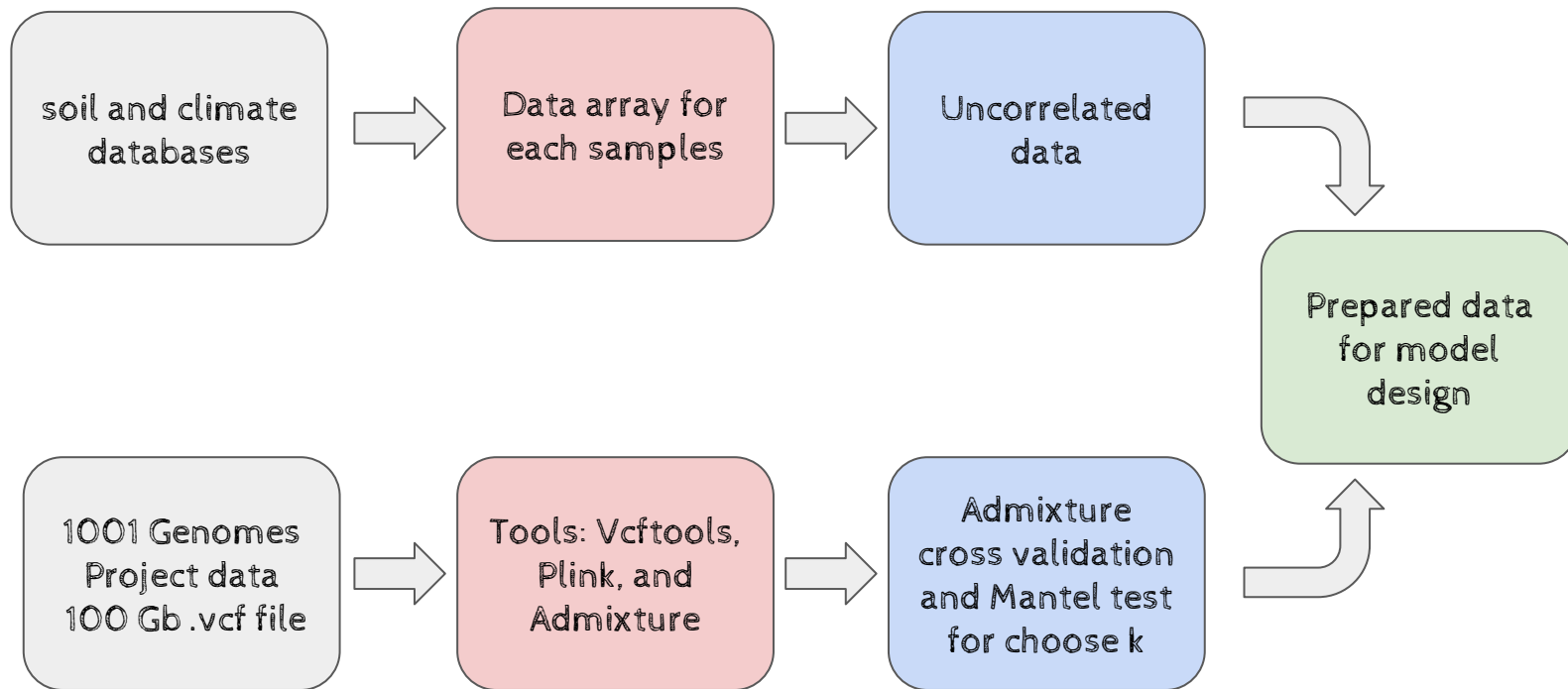
1. Determine the climate and soil influence on genome of modeling plants.
2. Build a model base on genome and environment data.



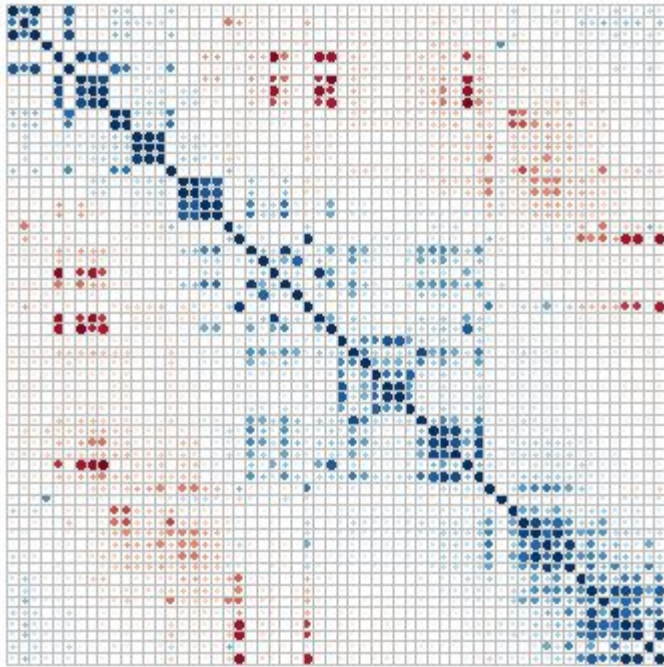
Tasks

- Get familiar with WorldClim and Soil databases
- Download 1001 genome SNPs data and collection coordinates of arabidopsis
- Get climate and soil data associated with this points
- Convert SNPs to vectors by Admixture tools
- Design model
- Interpretation of result

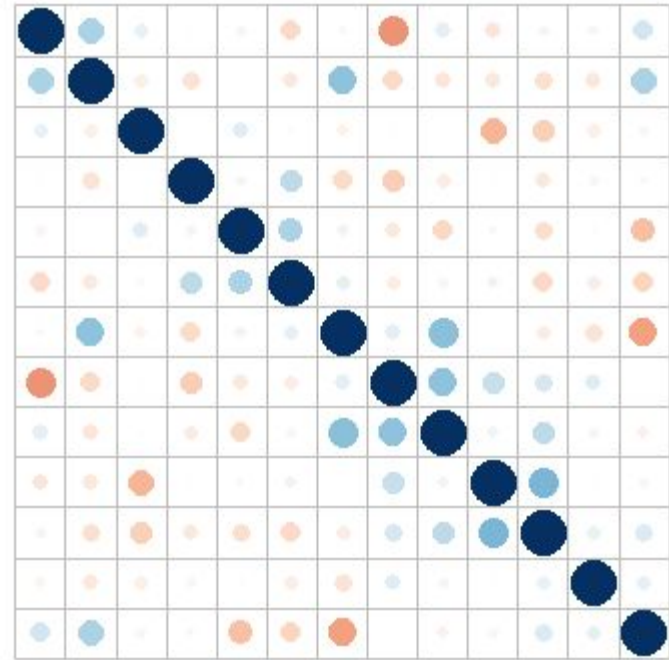
Roadmap of data preprocessing



Correlation

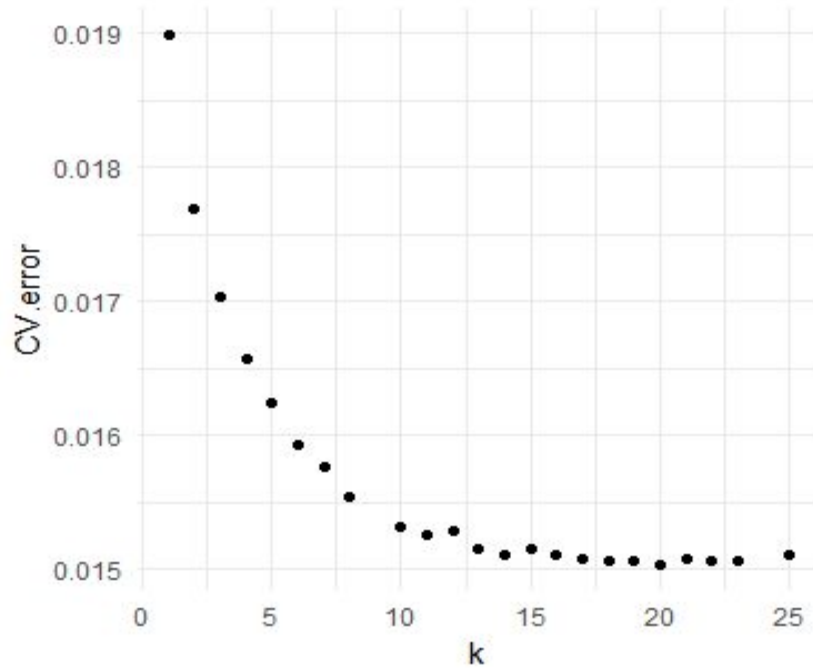


Correlation plot for initial data
(58 vars)



Correlation plot cutoff = 0.5
(13 vars)

Admixture cross validation and Mantel test



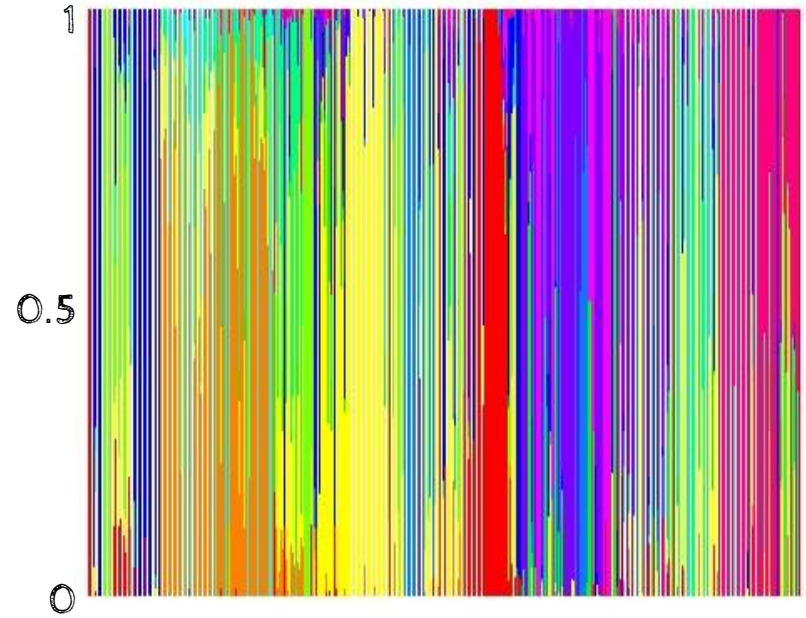
Cross validation plot

k	Observation	p-value
2	0.3520889	0.03225806
3	0.6759539	0.03225806
5	0.6897299	0.03225806
12	0.8257159	0.03225806
20	0.8384786	0.03225806

Admixture components

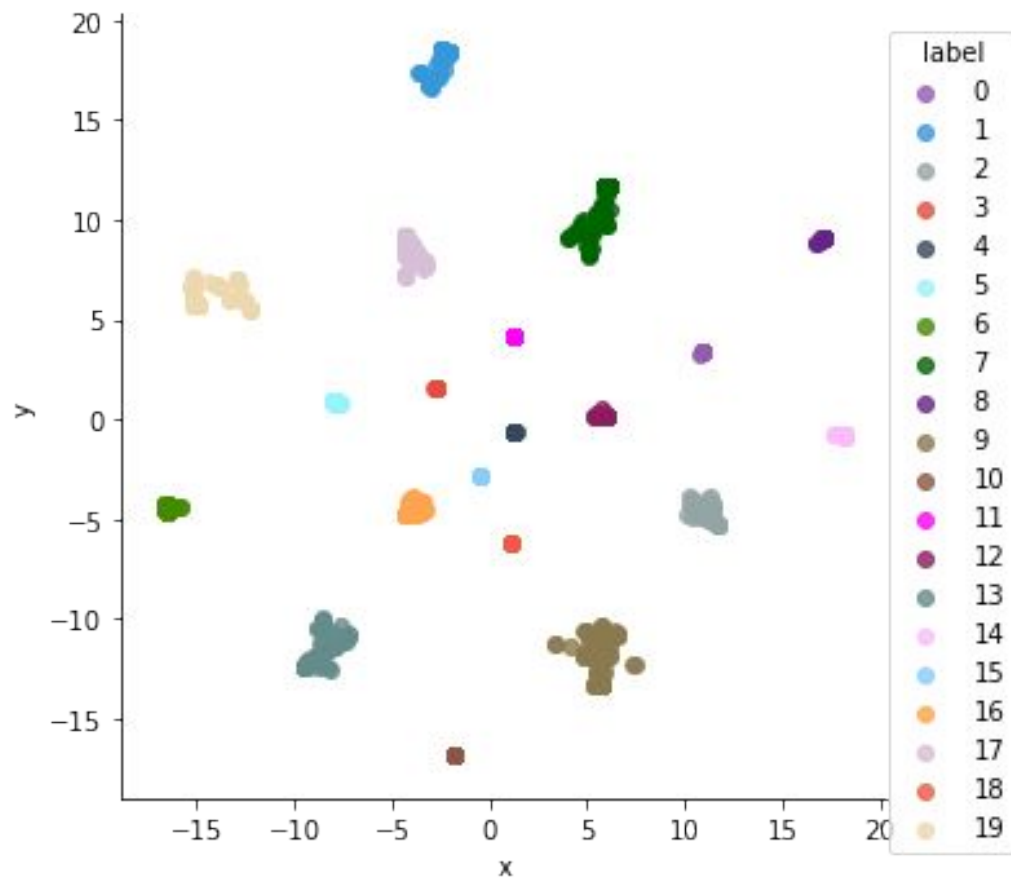


unordered samples $k = 20$



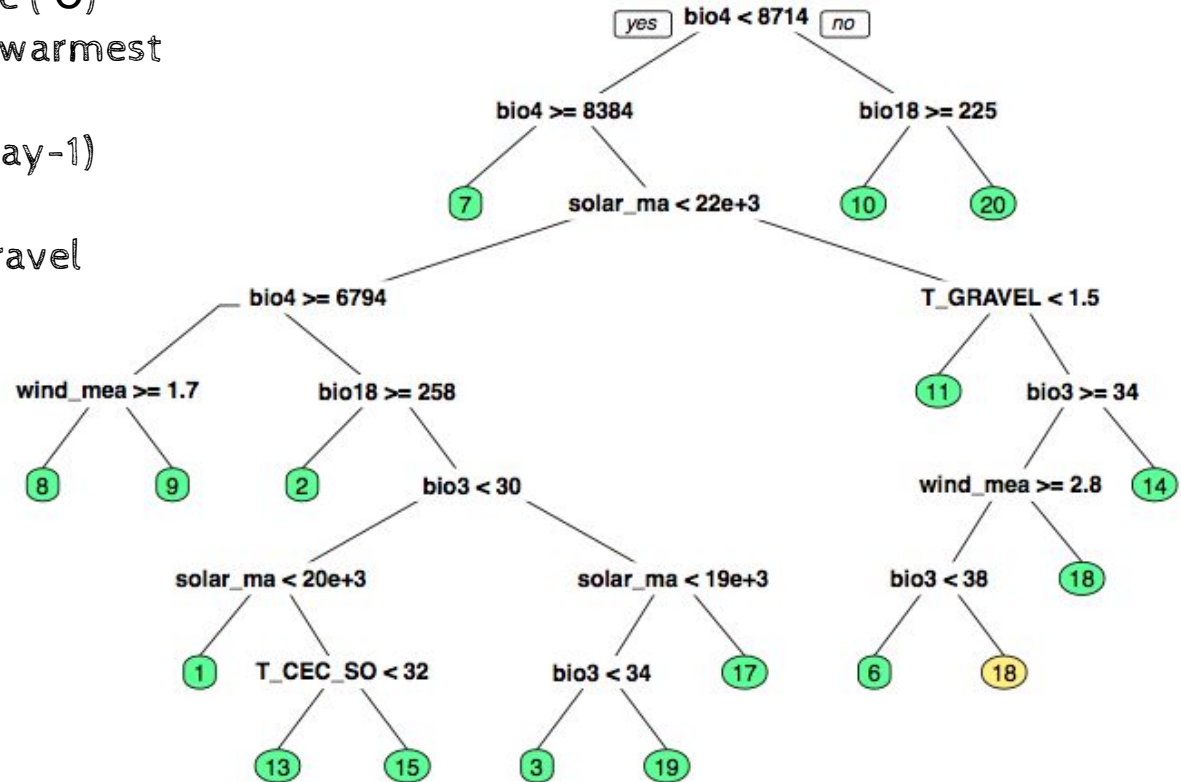
samples grouped by country $k = 20$

t-SNE



Regression tree

1. bio3, bio4 - temperature (°C)
2. bio18- precipitation of warmest quarter
3. solar radiation(kj m-2 day-1)
4. wind speed(m/s)
5. T_GRAVEL - Ttopsoil Gravel Content
6. T_CEC_SOIL - Topsoil CEC (soil)



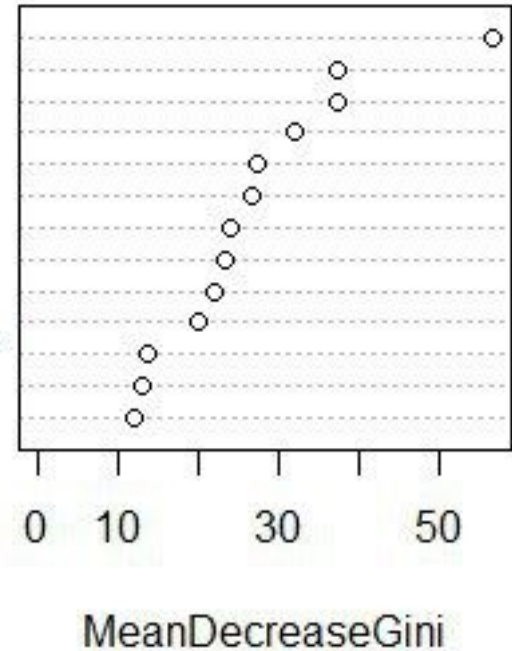
Random forest

1. bio4, bio8 - temperature
2. bio18, bio13, bio15-
precipitation
3. Water vapor pressure
4. wind speed
5. T_GRAVEL - Gravel Content

6. T_OC - Organic Carbon
7. T_ESP - Sodicity (ESP)
8. T_SAND - Sand Fraction
9. S_BS - Base Saturation
10. S_CEC_CLAY - CEC (clay)

bio4
vapr_mean
bio18
wind_mean
bio8
bio13
bio15
T_GRAVEL
T_OC
S_BS
S_CEC_CLAY
T_ESP
T_SAND

res.forest



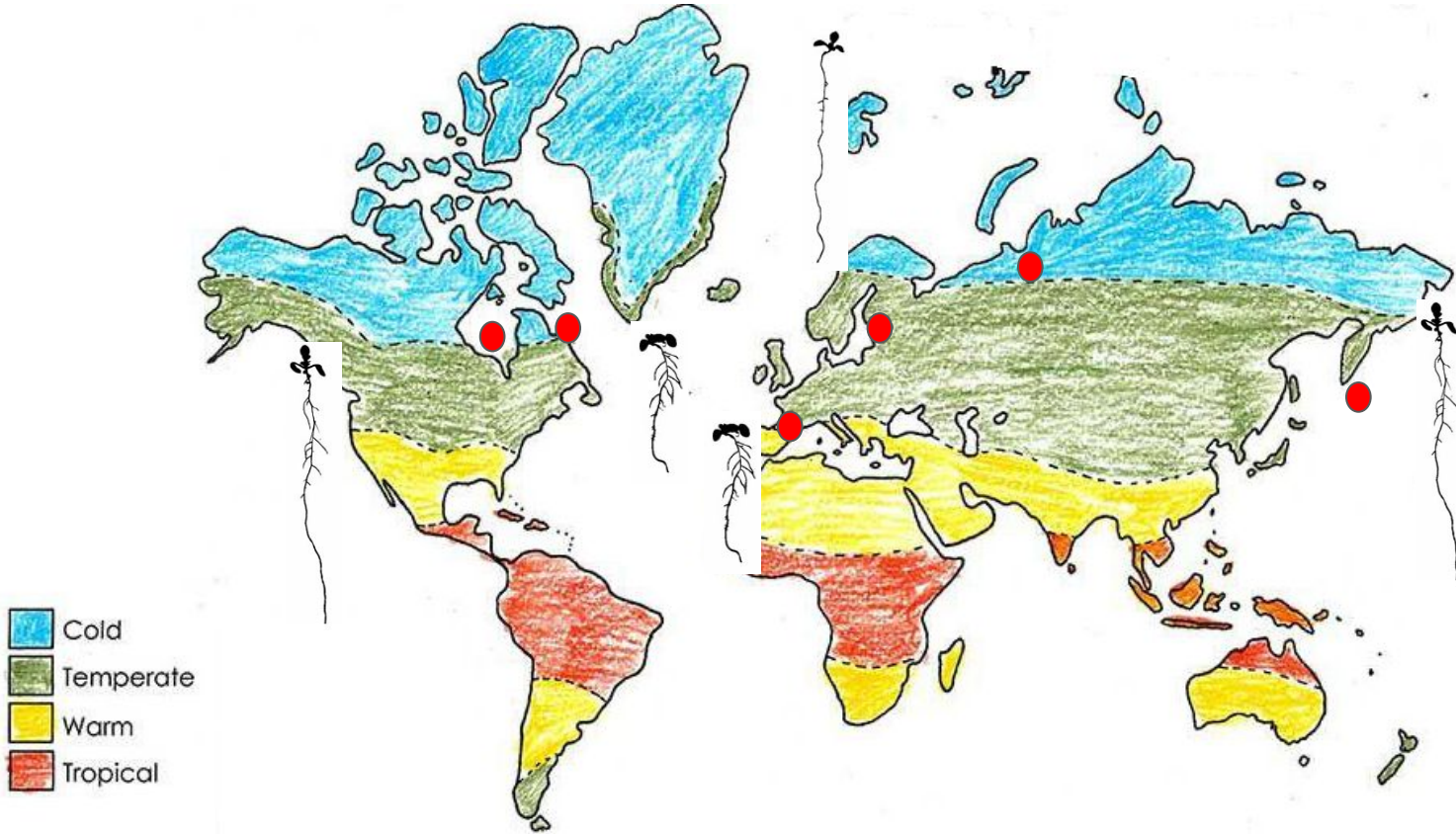
Results and discussion

1. Scripts for extracting climatic and soil data
2. Design models of predictions admixture class of *arabidopsis*
3. The parameters explaining the division of *arabidopsis* into admixture classes. It is found out what is most important for accurate prediction climatic variable which is consistent with the recent results.

Stearns F. W., Fenster C. B. Evidence for parallel adaptation to climate across the natural range of *Arabidopsis thaliana* //Ecology and evolution. – 2013. – T. 3. – №. 7. – C. 2241-2250.

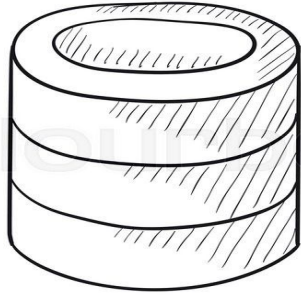
4. Git [repository](#)

Thank you for attention



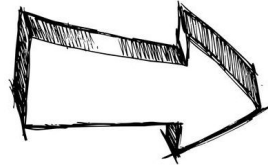
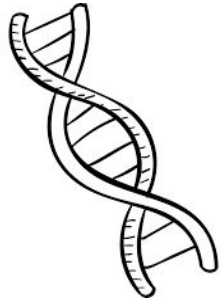
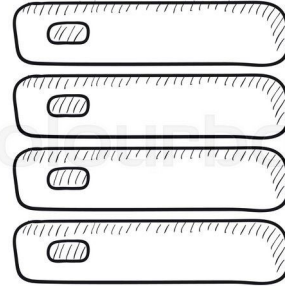
Genome data to matrix

1001 Genomes Project data
100 Gb .vcf file



Run tools on
the server

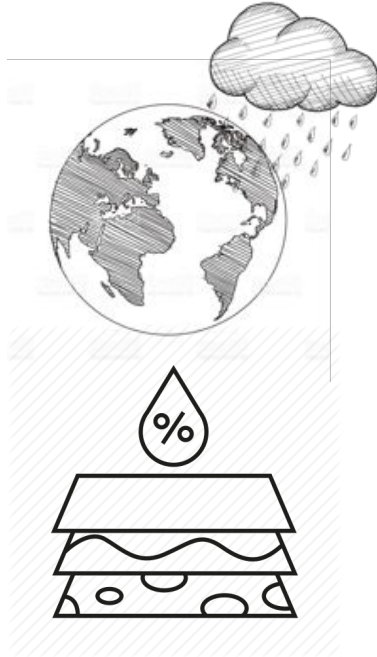
Tools: Vcftools,
Plink, and
Admixture



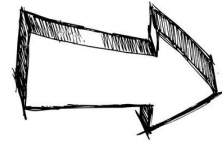
1. 0. 0. ... 0.
.9 .1 0. ... 0.
...
0. 1. 0. ... 0.

Climate and soil databases

WorldClim and Soil data



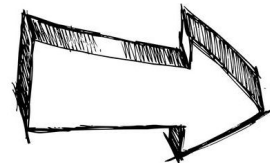
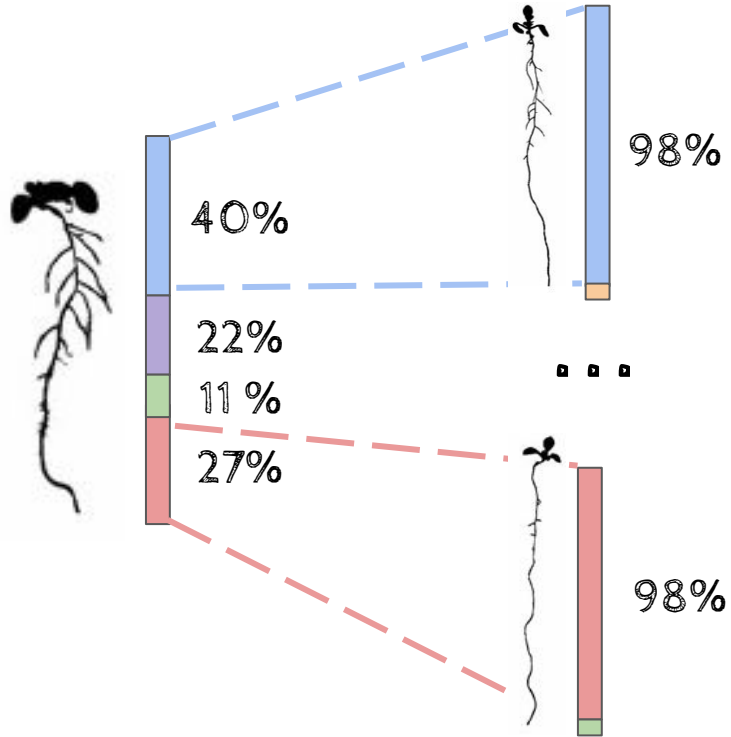
R script



parameters for
all given points

$$\begin{matrix} b_{00} & b_{01} & \dots & b_{0m} \\ b_{10} & b_{11} & \dots & b_{1m} \\ & & \dots & \\ b_{n0} & b_{n1} & \dots & b_{nm} \end{matrix}$$

Admixture



- 0.4 a_{00}
- 0.22 a_{01}
- ...
- 0.27 a_{0k}