

BIOINFORMATICS
INSTITUTE

Phylogenetic Networks Comparison

Supervisor:

Alekseev Nikita,
ITMO University

Students:

Eliseev Anton

Klimenko Natalya

Pazhenkova Elena

The problem

Sequence data => phylogenetic trees from
different markers => networks

All networks have the same number of reticulation
events.

How to compare them?

Idea:

use the number of possible convex colorings as
metrics of network's quality

Challenges

Data preprocessing

- formats of data
- detect outliers

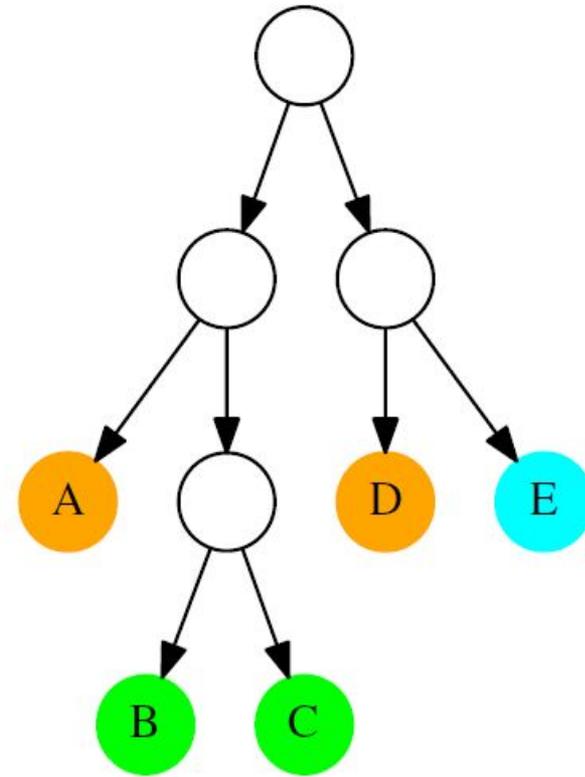
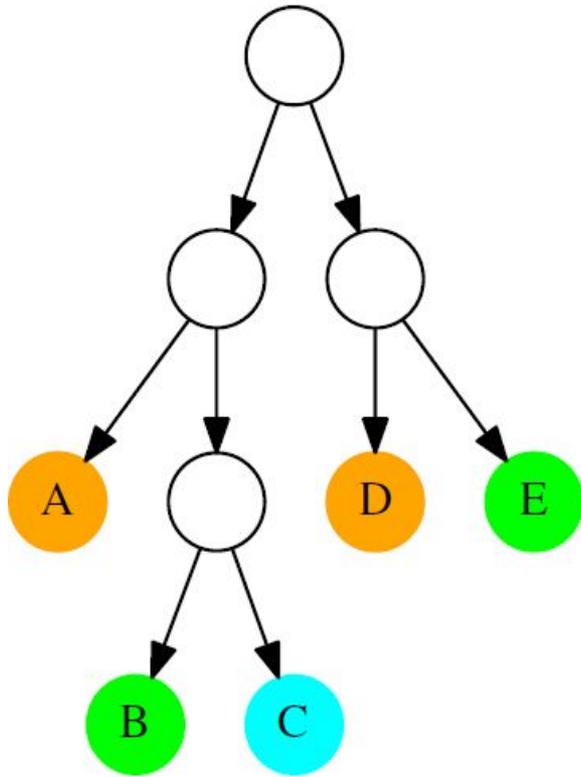
Tools

- Dendroscope, SplitsTree
- Extended Newick (2008)

Algorithm (clever bruteforce):

- Counting of convex colorings

Convex coloring vs homoplasy



Heliconius butterflies

- Traditional example of Müllerian mimicry and object for the study of convergence.
- At least 27% of all species produce hybrids in the wild.



melpomene melpomene



heurippa



cydno cordula



melpomene ssp

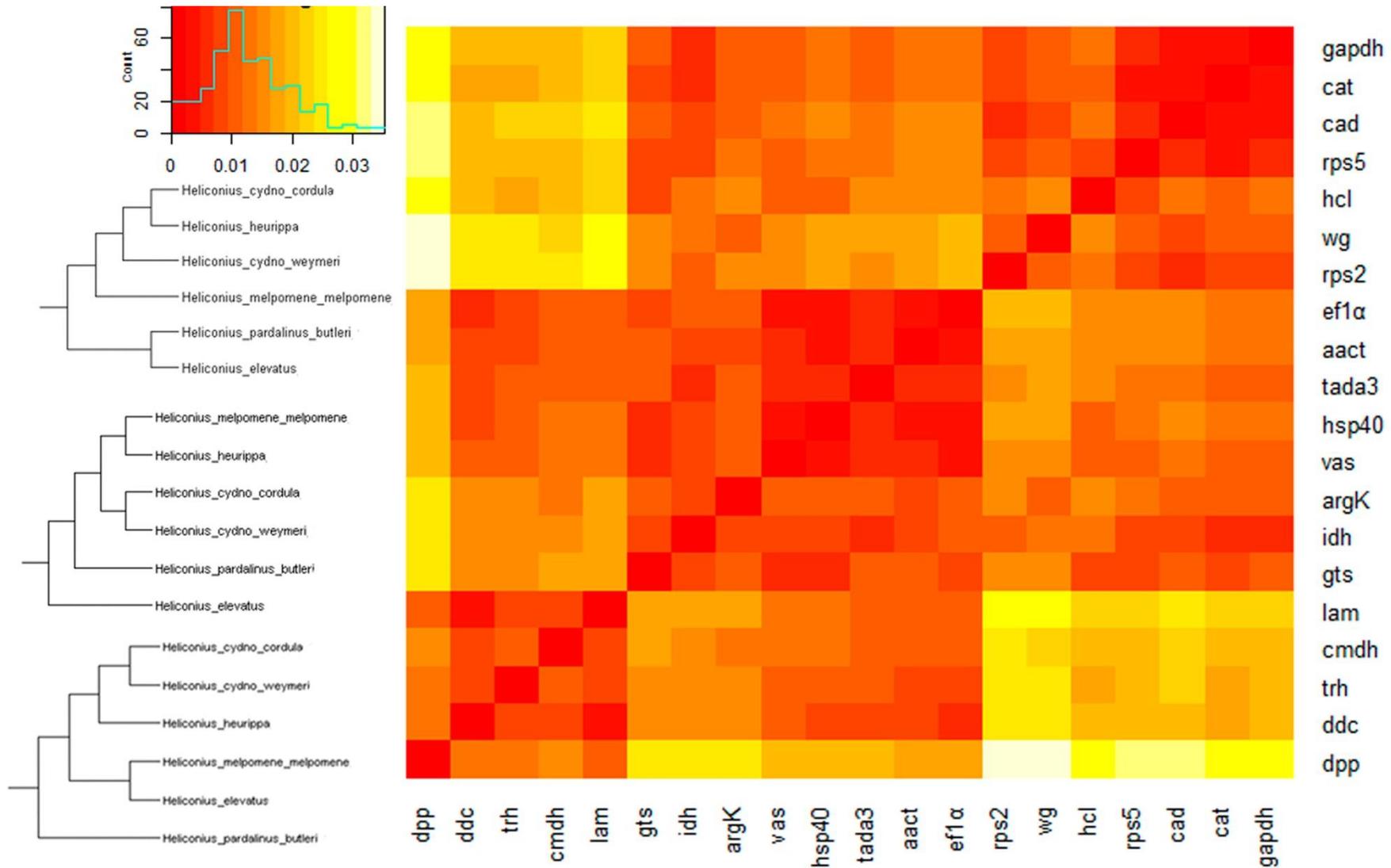


elevatus



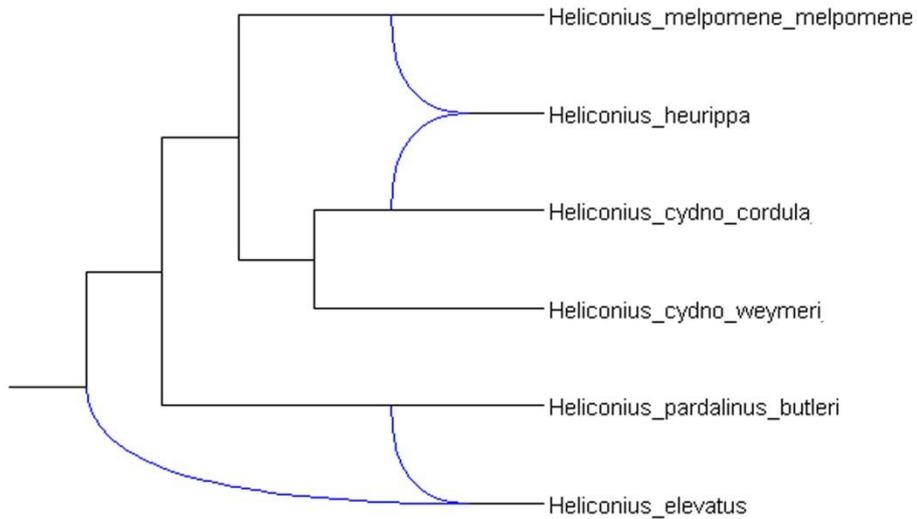
pardalinus butleri

Analysis of 20 protein-coding nuclear genes

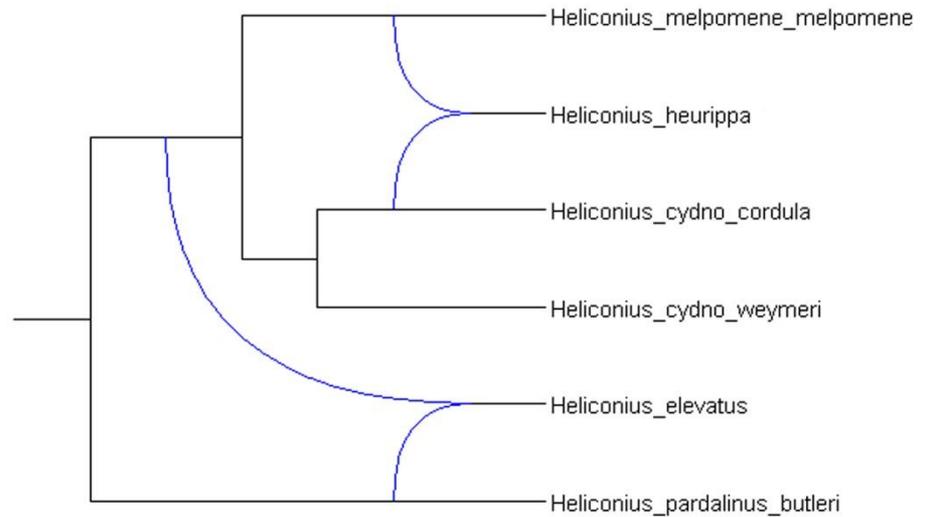


Heatmap of pairwise Branch Score Distances (Kuhner & Felsenstein, 1994), estimated in R (ape package).

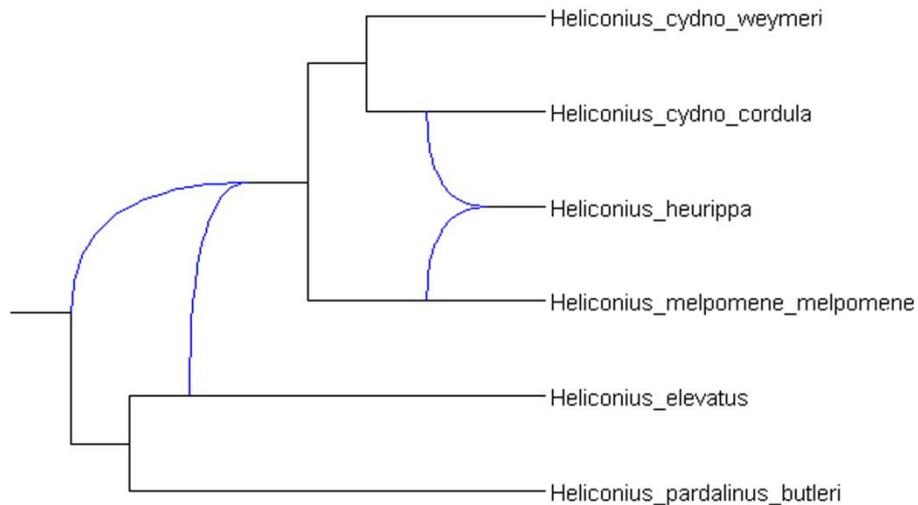
Hybridization networks



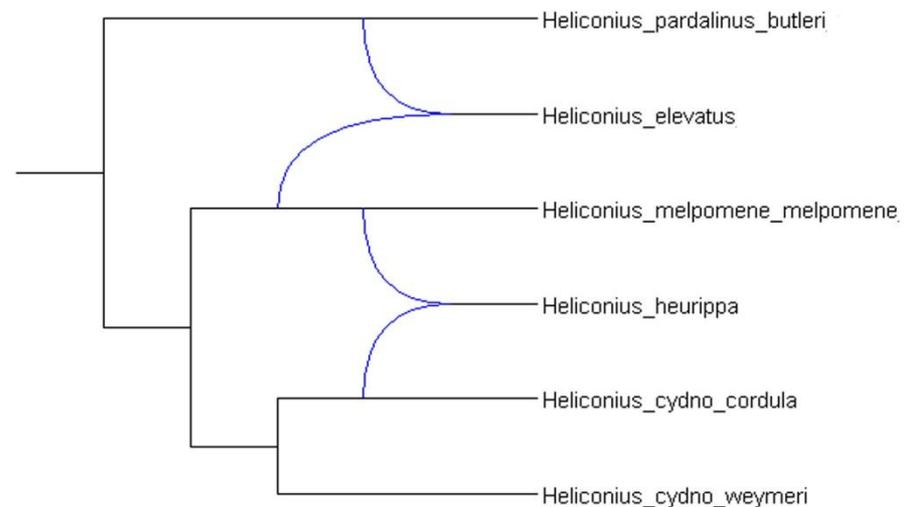
106



106



106



120

POTATOES



CIP

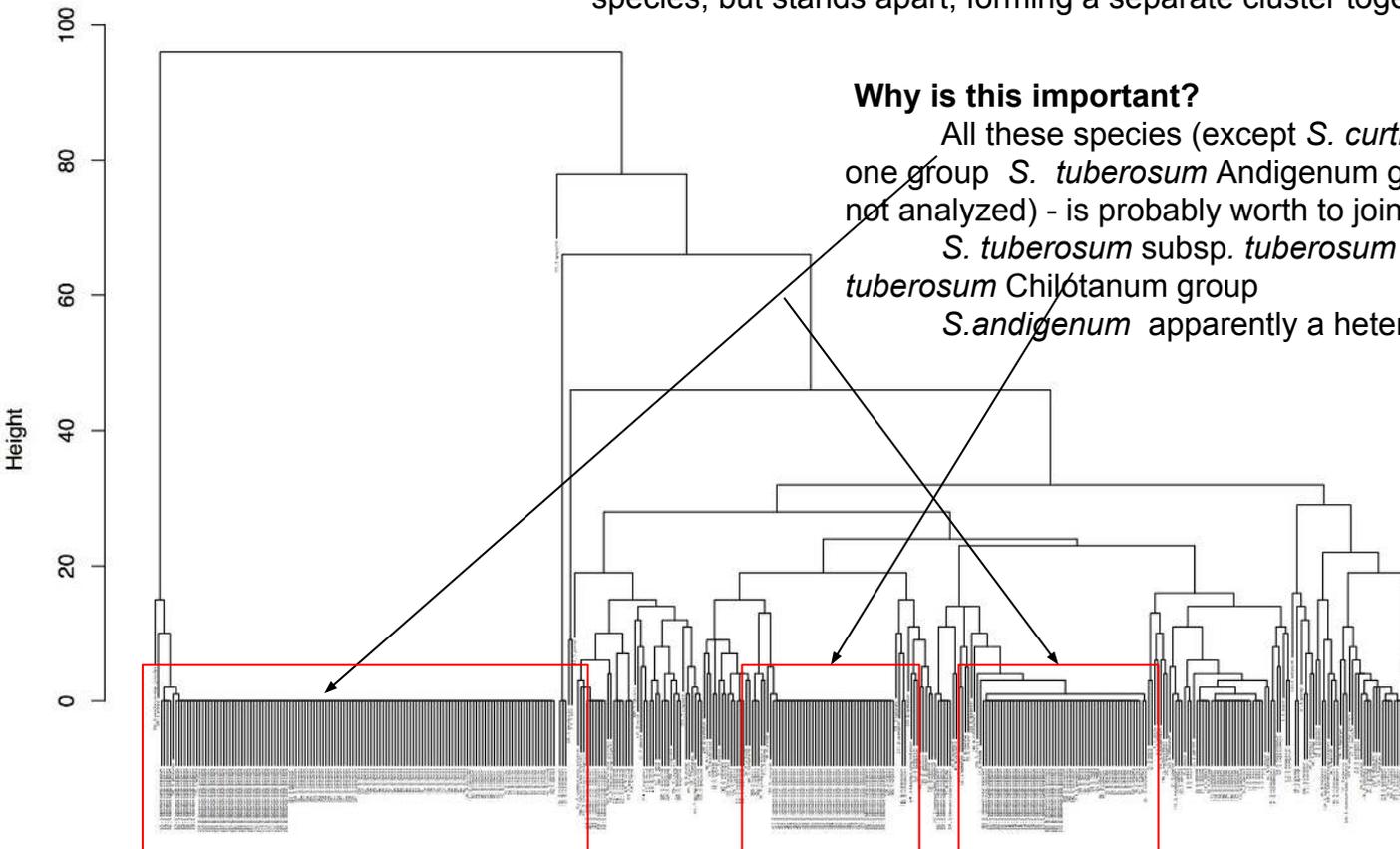
Important points:

- 1) All potatoes are divided into wild (just grow in nature) and cultivated (widely used in breeding, cultivated by the local population of South America) species
- 2) There are several classifications of potato species
 - according to S.M. Bukasov, 1972 (more than 170 species, 11 of them are cultivated ones)
 - Hawkes, 1990 (235 species, 7 cultivated)
 - Spooner et. al, 2014 (111 species, 4 cultivated), created on the basis of anatomical, morphological, molecular data — RAPD, ISSR, nSSR, AFLP, CAPS, STS methods. All previous classifications were based only on anatomical and morphological data.

Dendrogram of 420 samples of 29 species

Interesting moments:

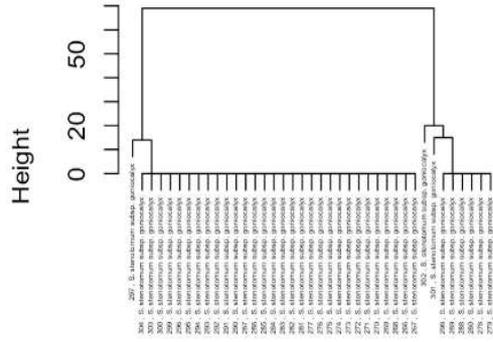
Solanum stenotomum subsp. *goniocalyx*, *S. stenotomum*, *S. phureja*, *S. chaucha*, *S. curtilobum* - form two large clusters. These two clusters also include 30 of 50 samples *S. andigenum*.
S. tuberosum subsp. *tuberosum* is not included in any of the two large clusters of cultivated species, but stands apart, forming a separate cluster together with samples *S. andigenum*.



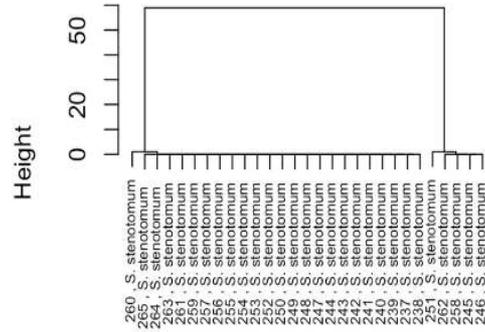
The result of the analysis of the plastid microsatellite sequences using 15 types of SSR-markers

Solanum tuberosum Andigenum group

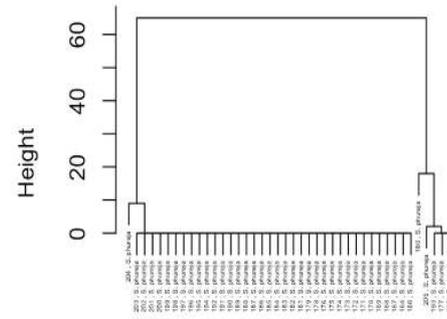
S. stenotomum subsp. *goniocalyx*



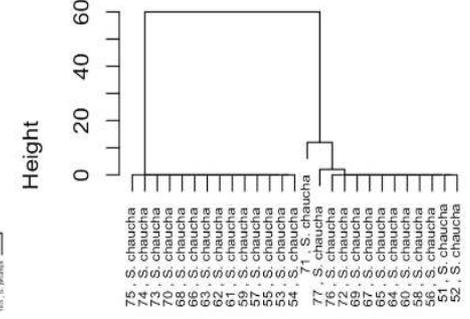
S. stenotomum



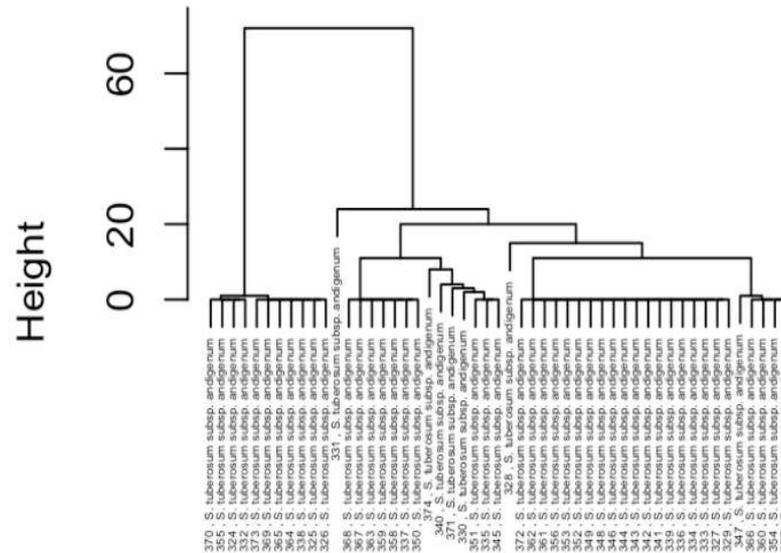
S. phureja



S. chaucha

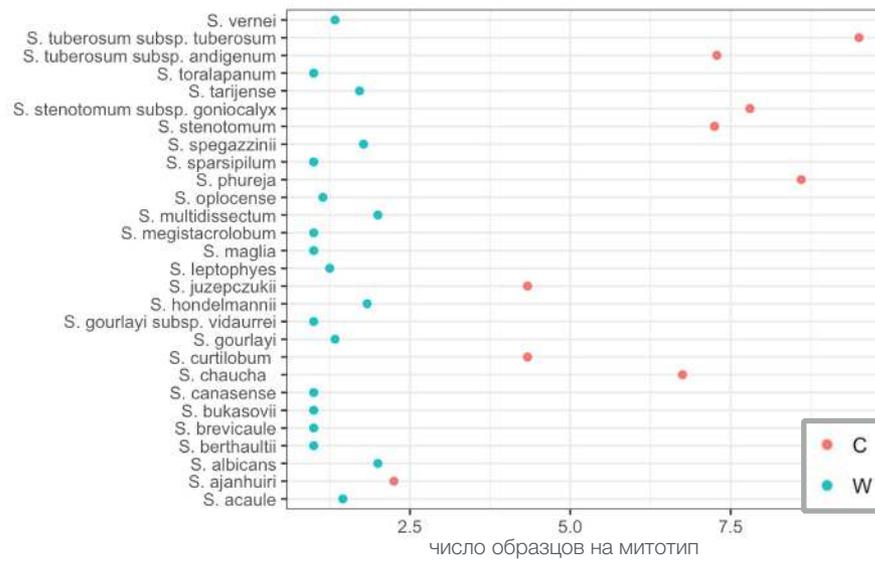


S. andigenum



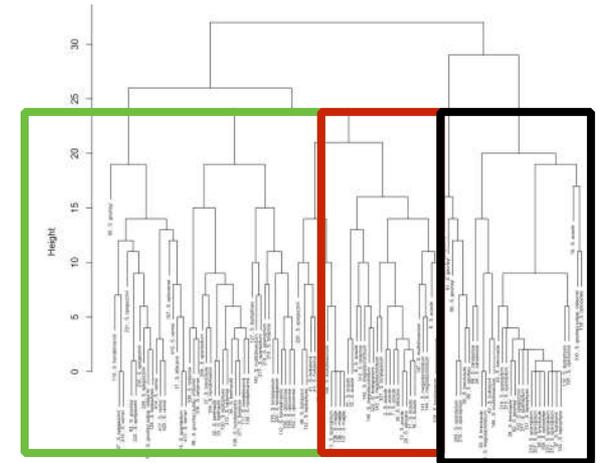
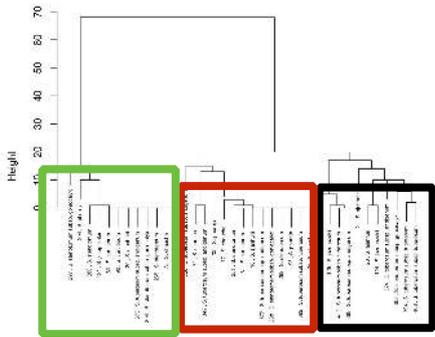
For all other species, individual dendrograms are made too

Samples with a minimum distance between each other in the matrix (manhattan distance method) were further considered as one sample.



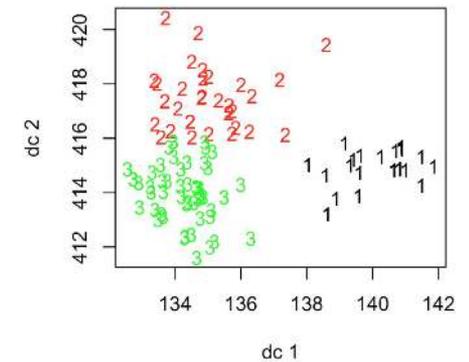
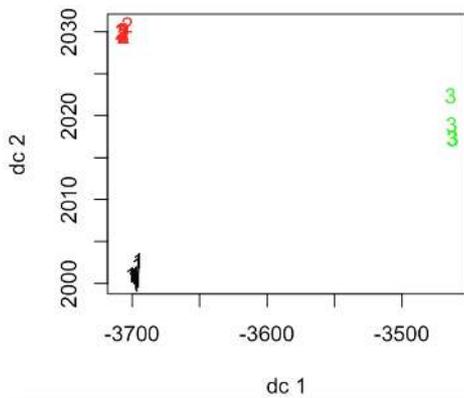
Dendrogram of cultivated potato species

Dendrogram of cultivated potato species

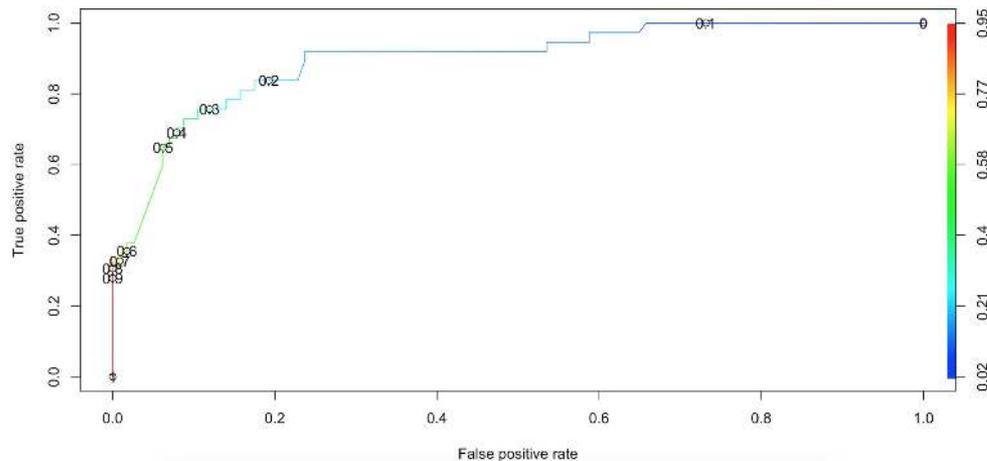


All cultural species are intermixed. As a result, they can be divided into 3 autonomous groups.

All cultural species are intermixed. As a result, they can be divided into 3 groups.



Analysis of the sample's belonging to the group of cultivated / wild species



The AUC ROC value (Area Under Curve) was 0.8974.

Conclusion

Cultivated and wild species of potato are clearly distinguishable among themselves

Probably a classification based on the anatomical and morphological features does not reflect the diversity of species at the molecular level

the idea of dividing cultural species into *Solanum tuberosum* Andigenum group and *Solanum tuberosum* Chilotanum groups - is correct

Thanks!

