

# Selection Events in *Allium* (Amaryllidaceae) Plastomes

**Speaker:** Pavel Nikitin, Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University

**Authors:** Scobeyeva V.A.<sup>1,2</sup>, Artyushin I.V.<sup>1</sup>, Krinitsina A.A.<sup>1</sup>, Nikitin P.A.<sup>1\*</sup>, Antipin M.I.<sup>1</sup>, Kuptsov S.V.<sup>1</sup>, Belenikin M.S.<sup>2</sup>, Omelchenko D.O.<sup>3</sup>, Logacheva M.D.<sup>4</sup>, Konorov E.A.<sup>5</sup>, Samoilov A.E.<sup>6</sup>, Speranskaya A.S.<sup>6,1</sup>

<sup>1</sup> Lomonosov Moscow State University, Moscow, Russia

<sup>2</sup> Moscow Institute of Physics and Technology, Moscow, Russia

<sup>3</sup> Institute for Information Transmission Problems (RAS), Moscow, Russia

<sup>4</sup> Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia

<sup>5</sup> Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

<sup>6</sup> Central Research Institute of Epidemiology (CRIE), Russia

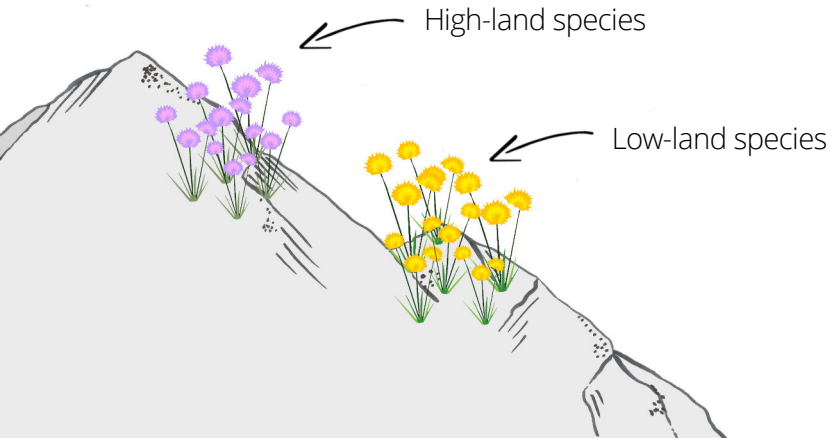
\* email: [nikitinp@fbb.msu.ru](mailto:nikitinp@fbb.msu.ru)

# Introduction

Amaryllidaceae is a large family with more than 1600 species, belonging to 75 genera. The largest genus - *Allium* comprises about 1000 species. They are widespread and are adapted to a wide range of habitats. They can even live on mountains at an altitude of 5000 meters. Genes present in chloroplast genomes (plastomes) play fundamental role for the photosynthesis. Plastome traits could thus be associated with geophysical abiotic characteristics of habitats. Most chloroplast genes are highly conserved and are used as phylogenetic markers for many families of vascular plants. Nevertheless some studies revealed signatures of positive selection in chloroplast genes of many plant families including Amaryllidaceae. In this work we provide analysis of Allioideae subfamily plastid genomes selection events.



Figure 1. *Nothoscordum* sp.  
reference: wikipedia.org



## Aims:

1. To detect selection events among *Allium* genus plastomes.
2. To infer if selection is acting on plastome genes in different habitat altitudes.

# Pseudogenization

- MAFFT (multiple alignment using fast Fourier transform) – primary alignment
- MACSE (Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons) for pseudogene search
- IQ-Tree2 (Maximum Likelihood phylogenetic reconstruction)

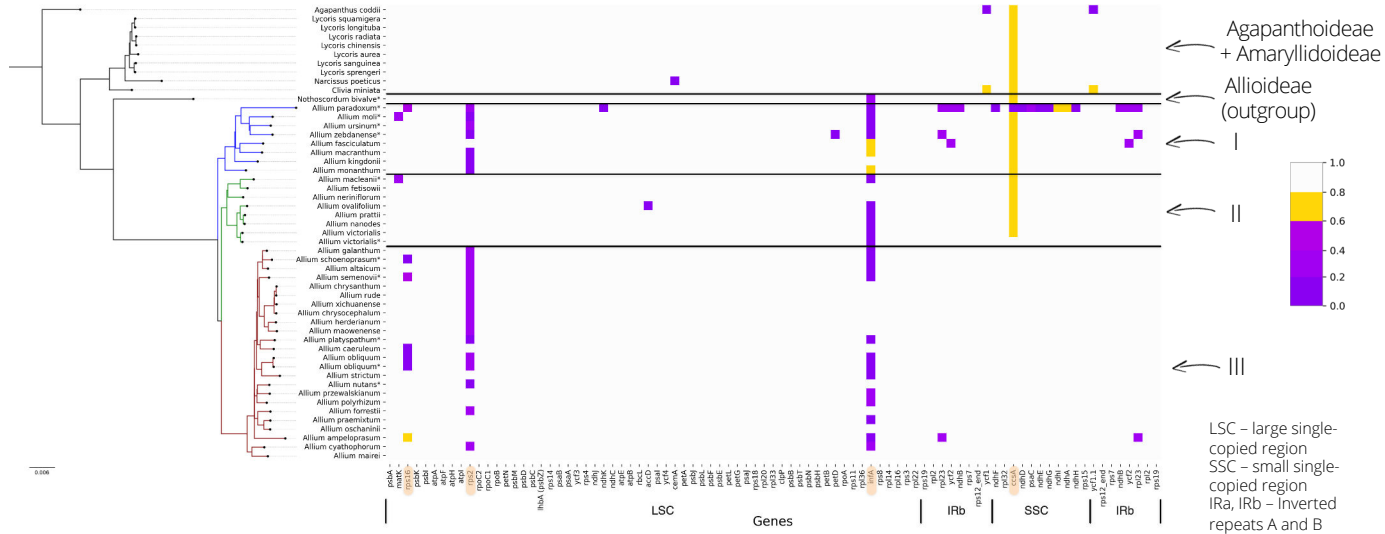


Figure 2. Maximum likelihood phylogenetic tree and pseudogenization heatmap in the plastome of Amaryllidaceae (including all three evolution lines of *Allium*). Color shows a proportion of remaining gene product.

# Evolution rates along the genome



Proportion of gaps in alignment in the window of 100 b.p.

GC-content in the window of 100 b.p.

Relative tree length in the window of 100 b.p.

Relative mutation rate barplot

$\frac{dN}{dS}$ -MEME rate barplot

Gene GC-content

Figure 2. Information on alignment and gene features.

# Selection and environmental conditions



To detect selection signals we used different methods from HyPhy framework (Hypothesis Testing using Phylogenies):

- aBSREL (adaptive Branch-Site Random Effects Likelihood)
- FUBAR (Fast Unconstrained Bayesian AppRoximation)
- MEME (Mixed Effects Model of Evolution)

Evolutionary line	Species	Habitat altitude (meters)	Genes
I	<i>Allium paradoxum</i>	No data (lives in shady forests)	<i>matK</i>
	<i>Allium macranthum</i>	2700 – 4200	<i>ndhF, rpl16, rpoC2</i>
	<i>Allium kingdonii</i>	4500 – 5000	<i>ndhK</i>
II	<i>Allium neriniflorum</i>	500 – 2000	<i>ndhF, rpoB</i>
III	<i>Allium strictum</i>	No data (lives on open rocks)	<i>ndhK</i>
	<i>Allium przewalskianum</i>	2000 – 4800	<i>ndhj</i>
	<i>Allium forrestii</i>	2700 – 4200	<i>ndhF</i>
	<i>Allium oschanii</i>	3000	<i>ndhj, rpoB</i>
	<i>Allium cyathophorum</i>	2700 – 4600	<i>ndhD, ndhj</i>

Table 1. Species in which different methods showed positive selection

# Conclusion

- Genes *infA*, *ccsA*, *rps2* and *rps16* have lost their functionality multiple times in different species, while the pseudogenization of other genes was occasional.
- The “normal” or “pseudo” state of *rps2* and *ccsA* genes correlates well with the evolutionary line of genus the species belong to.
- Independent methods revealed some housekeeping genes (*accD*, *matK*, *rpoB*), photosynthesis-involved genes (*ndhD*, *ndhF*, *ndhK*, *rbcL*) and genes of unknown function (*ycf1*, *ycf2*) being under positive selection.
- Most species in which genes are being under positive selection live high in the mountains (more than 2000 meters above sea).
- Taking into account known mechanisms of coping with excessive light by cyclic electron transport, we can hypothesize that adaptive evolution in genes, coding subunits of NADH-plastoquinone oxidoreductase could be driven by abiotic factors in high habitat altitudes like temperature, light intensity or UV radiation.

## Acknowledgements

I express gratitude to my tutors A. S. Speranskaya, A. A. Krinitsina, I. V. Artyushin, V. A. Scobeyeva, D. V. Pozdyshev and my mate E. N. Pitikov for guidance and support.



Custom scripts are available in GitHub repository:  
[https://github.com/nikitin-p/Allium\\_analysis](https://github.com/nikitin-p/Allium_analysis)