Microevolution Events in *Allium* (Amaryllidaceae) Plastomes

Pavel Nikitin (nikitinp@fbb.msu.ru)[1]
[1] Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Russia

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 from shady forests to meadows, steppes and deserts. Species 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.

Aims

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

Materials and methods

- cpDNA sequences (38 *Allium* species and 11 Amaryllidaceae species as an outgroup)
- 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)
- Biopython – ParsimonyScorer (estimation of substitution count for every window with a fixed tree topology)

HyPhy framework (Hypothesis Testing using Phylogenies):
- Relative tree length in alignment blocks
- $\frac{dN}{dS}$ rate for evaluating the balance between neutral mutations and positive/negative natural selection
- aBSREL (adaptive Branch-Site Random Effects Likelihood)
- FUBAR (Fast Unconstrained Bayesian AppRoximation)
- MEME (Mixed Effects Model of Evolution)

Plots were constructed with:
- Python3 pandas, numpy, seaborn and matplotlib packages
- R ggplot2 package

Custom scripts are available in GitHub repository:
https://github.com/nikitin-p/Allium_analysis

Results

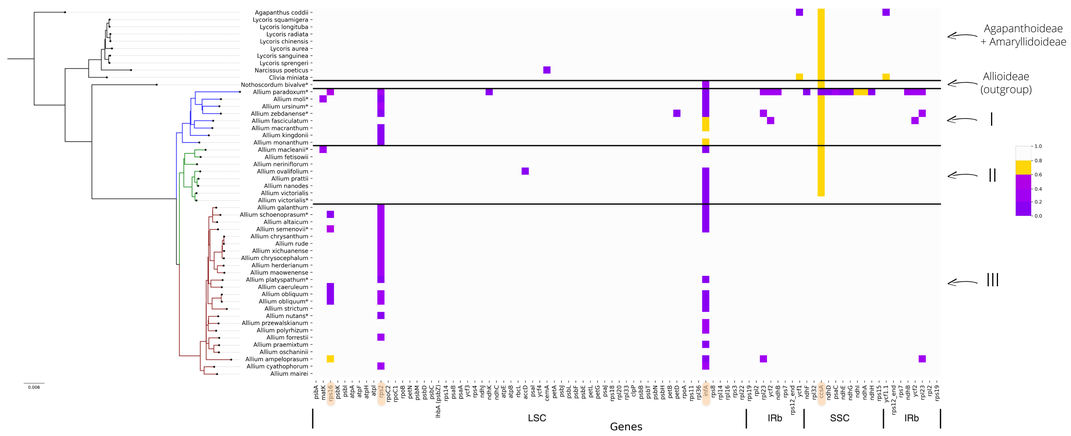

Figure 1. Maximum likelihood phylogenetic tree and pseudogenization heatmap in the plastome of Amaryllidaceae (including all three evolution lines of *Allium*). Color shows a proportion of remaining gene product.

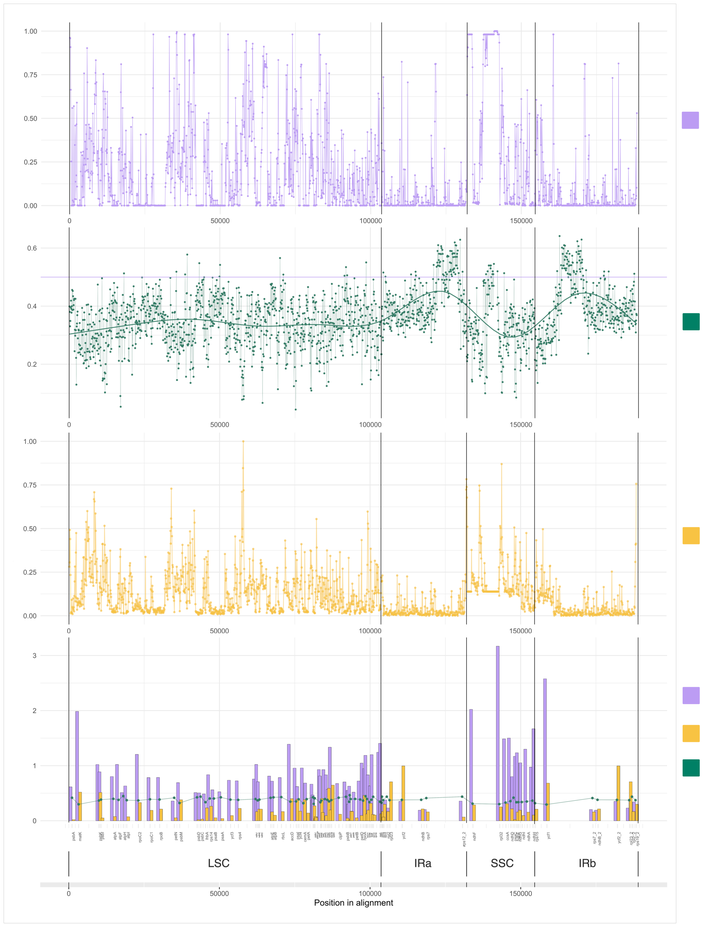

Figure 2. Information on alignment and gene features.

Table 1. Sites under positive or diversifying selection in protein coding genes.

Protein	Gene	Number of sites in alignment	Number of MEME sites under + selection	Number of FUBAR sites under + and – selection	Positions and number of MEME sites, confirmed by FUBAR
acetyl-CoA carboxylase carboxyltransferase beta subunit	*accD*	480	6	+5 -23	96, 156, 176, 477 total 4
Protein TIC 214	*ycf1*	1761	51	+9 -44	352, 705, 810, 851 total 4
Hypothetical chloroplast RF21	*ycf2*	2294	15	+35 -14	474, 475, 595, 691, 1786 total 5
NADH-plastoquinone oxidoreductase subunit 4	*ndhD*	507	5	+3 -88	404, 454 total 2
NADH-plastoquinone oxidoreductase subunit 5	*ndhF*	734	15	+6 -146	299, 510, 514,676 total 4
NADH-plastoquinone oxidoreductase subunit K	*ndhK*	249	3	+2 -20	235, 240 total 2
maturase K	*matK*	521	13	+7 -40	92, 324, 345 total 3
ribulose-1,5-bisphosphate carboxylase oxygenase large subunit	*rbcL*	480	9	+10 -70	91, 97, 225, 265 total 4
RNA polymerase beta subunit	*rpoB*	1071	9	+5 -158	7, 160, 1061 total 3

Table 2. Species in which aBSREL showed positive selection

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

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.