DOI 10.18699/SBB-2022-008

Longread-only approach to the organellar genome assembly of a rare endemic non-model species *Crepis callicephala* Juz. (Asteraceae)

Emirsaliiev A.[1, 2]*, Tsyupka V.[1, 2], Grebennikova O.[1, 2], Bulavin I.[1, 2], Krivenko O.[1, 2], Ivanova N.[2], Nikiforov A.[2], Mitrofanova I.[1, 2, 3]

[1] *Kurchatov Genomic Centre – NBG-NSC, Yalta, Russia*

[2] *FSFIS "Nikita Botanical Gardens – National Scientific Center of RAS", Yalta, Russia*

[3] *FSBIS Main Botanical Gardens named after N.V. Tsitsin, Moscow, Russia*

* *asantaraqtasli@gmail.com*

Key words: cpDNA, plastid genome, chloroplast, endangered species, adaptation, Crimea

Motivation and Aim: *Crepis callicephala* Juz. is a rare endemic species of Crimean flora. It belongs to the Asteraceae family – one of the most important for human and widespread eudicots family in the world. The *Crepis* genus had a great impact on the plant embriology and cytogenetics as a usefull object for chromosome observation [1]. But its genome has not been in detail investigated yet via molecular methods. The Crepidinae subtribe members of Asteraceae, including the genus *Crepis*, are the perspective source of different biologically active compounds of a high interest, as well some species of their parental tribe are included into local medicinal plant lists and pharmacopoeia. *C. callicephala* and relative species demonstrate notable niche specialisation [2] that also accords to the high level of endemism in this genus. Plastid genome codes a set of crucial photosynthetic genes [3] that contribute to the adaptive evolution. Being the wild relative of such important cultivated plants as *Lactuca* and *Cichorium*, the *Crepis* sp.sp. can be useful for plant geneticists to better understand adaptive abilities of that plants. As the modern omics approaches (including genomic researches) still has been unclear for the *Crepis* genus, we aimed to sequence and assemble the chloroplast genome of rare endemic Crimean species *C. callicephala*. The modern single molecule sequencing methods would be a good choice for that purposes due to relatively low cost and ability to resolve complex structure of plant organellar genomes in an assembly stage. Also we aim to compare different assembly methods in this application.

Methods and Algorithms: Seeds of *C. callicephala* was harvested previously in natural populations. *In vitro* plants were obtained by induction of multiple adventive proliferation from upper part of seedlings. DNA extraction, genomic library preparation and sequencing have been previously described [4]. Raw genomic reads base called with ONT Guppy v4.0.15 and v4.0.19, reads' statistics was scored using MinIONQC [5] NanoPack and LongQC packages. Adapters were trimmed with Porechop, reads were filtered using NanoFilt and bioawk.

For assembly, we initially selected the corresponding to plastids fraction of long reads. Reads have been aligned with minimap2. Unique mapped reads have been extracted using samtools.

Genome assembly was based on different reference-guided and *de novo* approaches. We used *de novo* tools such as Canu, flye, pomoxis (*de novo*) and NECAT [5], and also we

used pomoxis in a reference-guided mode. Assemblies were polished with long reads only using medaka. The final annotation was performed via GeSeq service [6] with enabling the CDS, tRNA and rRNA annotations on BLAT v36x7 (with search against Crepidinae and several other Cichorieae plastomes), HMMER v3.3.1, Chloë v0.1.0 and tRNAscan-SE v2.0.7.

Results: To examine the quality of selection previously extracted reads was one more realigned resulting in a good alignment with 94.34 % long reads mapped and total coverage depth x1,354.7. The longest mapped read was 124,313 bp in length with average 3,059 bp. Different assembly pipelines were performed. Some of them produced a single contig of expected length (NECAT and pomoxis in reference-guided mode), but the others (Canu, flye, pomoxis (*de novo*)) resulted either with very short or oversize and fragmented assemblies. Produced by NECAT and pomoxis assemblies had different total lengths, inverted repeat sizes and ycf1 gene sizes. To avoid such bias, we selected reads contained homologous to plastome sequences one more again. New reference was constructed from three plastomes, each of them was linearly concatenated – the one belonged to *L. sativa*, and the others were produced by pomoxis-medaka pipeline both in reference-guided and *de novo* modes. Mapped reads were used for assembly with pomoxis-medaka on *L. sativa* reference. Resulting plastome of *C. callicephala* had length 149,463 bp and contained 35 tRNA, 120 protein-coding genes and 13 putative pseudogenes.

Conclusion: Draft plastome sequence of rare endemic Crimean species *C. callicephala* Juz. was obtained from total genomic nanopore reads based on longread-only assembly approach. None of widely used *de novo* assembly tools worked well for this purpose.

Acknowledgements: The research was done within the framework of the grant assignment No. 075-15-2019-1670 and conducted on the base of the Unique Scientific Installation "Scientific Center of Plant Biotechnology, Genomics and Conservation" of FSFIS “NBG-NSC” and facilities of Shared Access Center Bioinformatics of FRC Institute of Cytology and Genetics of Siberian Branch of the RAS.

References

1. Navashin M.S. Karyology and Cytogenetics problems in Crepis genus research. Moscow: Nauka, 1985.
2. Nikiforov A.R. Aging population repertoire features of *Lagoseris callicephala* Juz. (Asteraceae). *Bull. State Nikita Botan Gard*. 2017;125:83-87.
3. Skobeyeva V., Omelchenko D., Logacheva M. et al. Plastid genome evolution in the genus *Allium*. In: Bioinformatics of genome regulation and structure/systems biology (BGRS/SB-2020): The Twelfth International Multiconference. Abstracts. Novosibirsk, 2020;243. doi: 10.18699/BGRS/SB-2020-154.
4. Emirsaliiev A.O., Brailko V.A., Grebennikova O.A. et al. Genomic research of the rare relict endemic plant *Crepis callicephala* Juz. (Asteraceae). In: Proceedings of the First All-Russian scientific and practical conference with international participation «Genomics and modern biotechnologies in plant propagation, breeding and preservation» October, 27–31, 2020. Simferopol: ARIAL, 2020;73-74. doi: 10.47882/GENBIO.2020.52.77.030.
5. Chen Y., Nie F., Xie Sh.-Q. et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun*. 2021;12(1):1-10. doi: 10.1038/s41467-020-20236-7.
6. Tillich M., Lehwark P., Pellizzer T. et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucl Acids Res*. 2017;45:W6-W11. doi: 10.1093/nar/gkx391.