Long-read nanopore sequencing to determine the nuclear genome of amphibious liverwort *Riccia fluitans* (Ricciaceae)

Katarzyna Krawczyk, Łukasz Paukszto, Jakub Sawicki

Department of Botany and Nature Protection,
University of Warmia and Mazury in Olsztyn

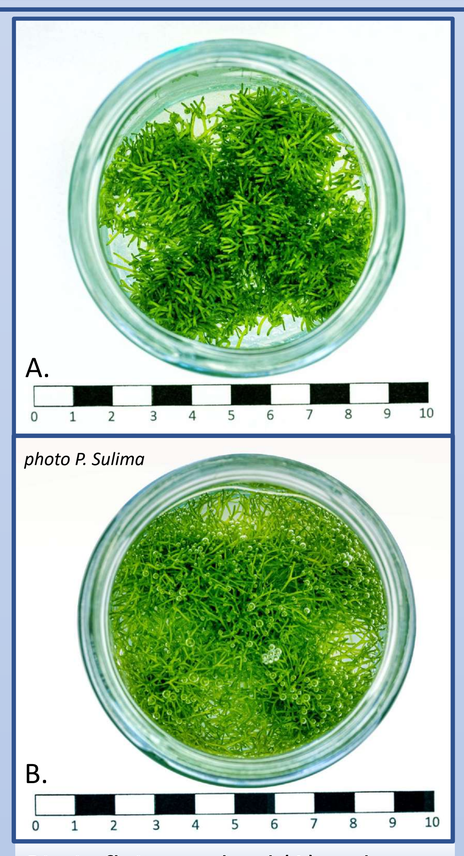
BACKGROUND

***Riccia fluitans* L.** (floating crystalwort) is a popular aquarium plant quickly growing and forming thick bright green mats on and under the water surface. It **can thrive in aquatic and terrestrial environments**, therefore in the wild it can be found both in water reservoirs and on denuded soil of cultivated fields or fallow lands.

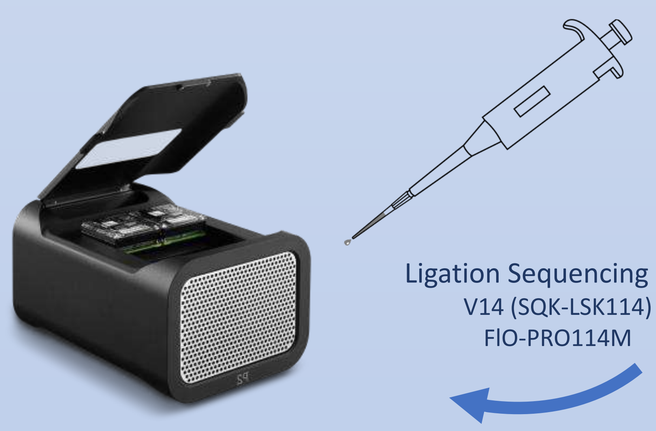
As a representative of Marchantiales, it belongs to the group of early-diverged land plants and due to its amphibious character, it makes an **excellent model for research on the adaptation of plants to the conquest of the terrestrial environment**. Understanding the structure of the nuclear genome is a necessary step for further research on this milestone evolutionary event.

Riccia fluitans – land and water form; *in vitro* culture

METHODS

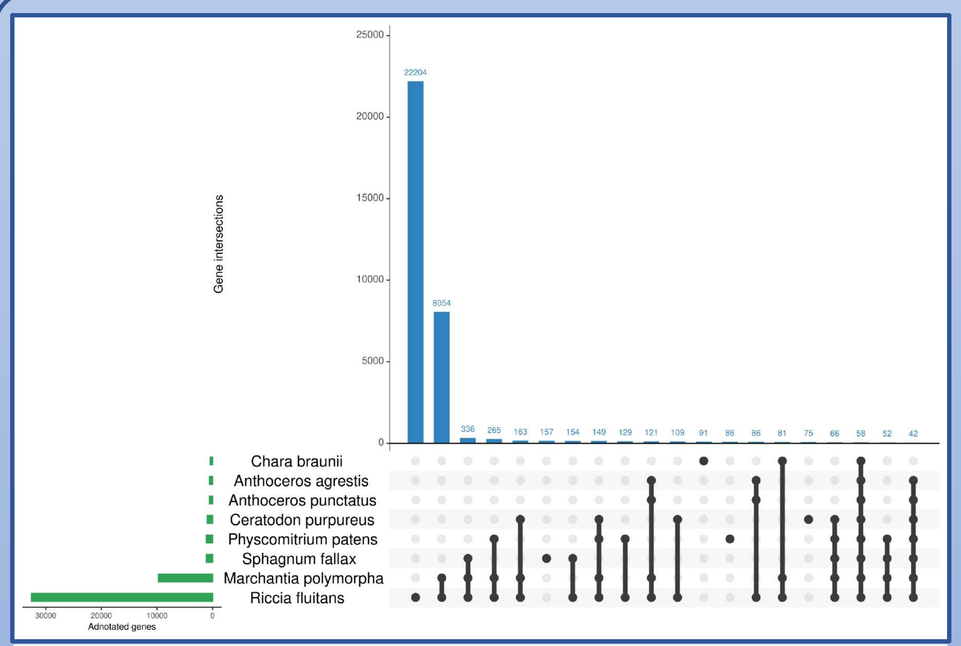


Riccia fluitans – land (A) and water (B) form – view from above

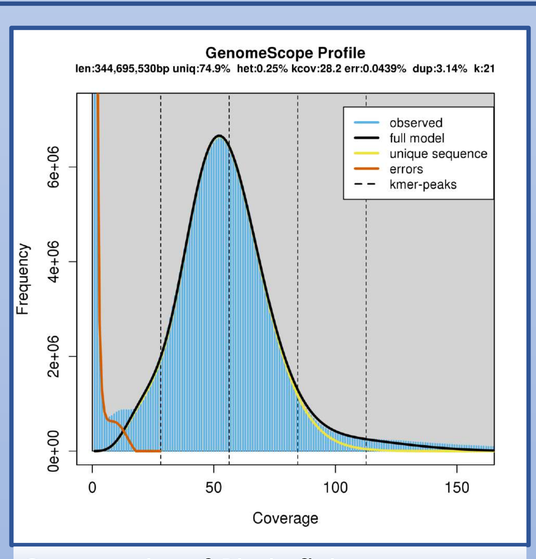


Super-accuracy parameters were used to basecall long Nanopore reads with **Dorado v0.1.1** software. To assembly contigs, **the flye 2.9.1** was applied with –meta flag and minimal overlap set to 1000 bp.
Next, contigs were annotated supporting by protein sequences of closely related liverwort *Marchantia polymorpha* (downloaded from marchantia.info). The exon-intron boundaries were predicted using **BRAKER3** software and co-programs **ProtHint**, **GENEMARK** and **AUGUSTUS**.
The completeness of the assembled genome and its annotation was tested by **BUSCO** software.

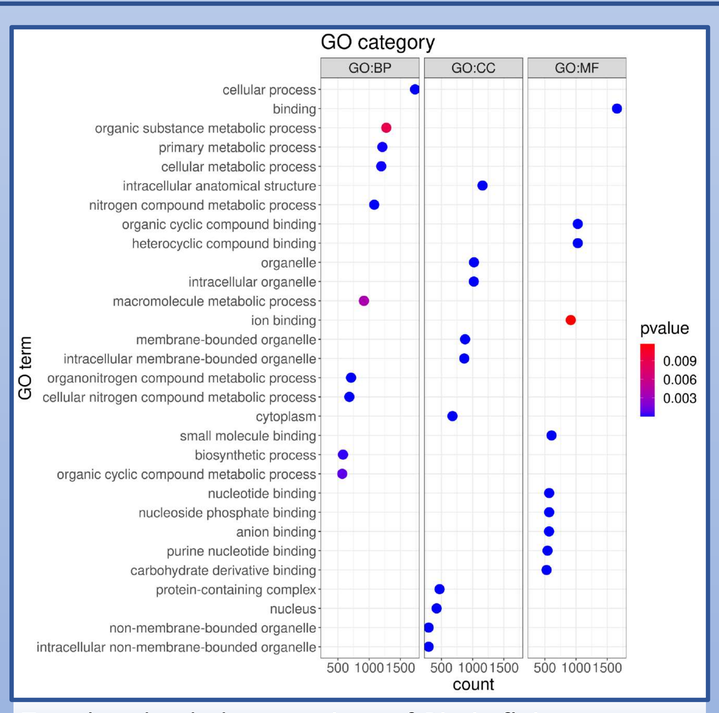
RESULTS & DISCUSSION



Ortholog analysis of eight plant species including *Riccia fluitans*.
The UpSet plot shows the overlap between 32,478 genes grouped into orthogroups from each species and the size of overlap as bar charts.



Genome size of *Riccia fluitans* genome estimated by GeneScope



Top abundanded annotations of *Riccia fluitans* genes according to *Marchantia polymorpha* homologous
GO- gene ontology; BP - biological process;
CC - cell component; MF - molecular function

Sequencing and genome assembly results

6600 contigs
470 Mbp
N50 = 5,324,768 bp
mean coverage = 131
the longest contig: 21,693,577 bp
146 contigs realign with 10 *Marchantia polymorpha* chromosomes

Completeness of the assembled genome and its annotation

405 (95,2%) complete BUSCOs:
- 381 (89,6%) complete and single-copy
- 24 (5,6%) complete and duplicated

10 (2,4%) fragmented BUSCOs
10 (2,4%) missing BUSCOs
425 (100%) total BUSCO groups searched

Characteristics of *Riccia fluitans* nuclear genome
A. gene frequency per 50k bp
B. number of LTRs per 50k bp
C. number of TIRs per 50k bp
D. read coverage – long Nanopore DNA seq.
E. read coverage – long Nanopore RNA seq.
F. short-read RNA coverage (>5000 in red)
G. 76 top-ranked expressed protein-coding genes

SUMMARY

- **95,2% of complete core Viridiplantae genes were detected;**
- **Only 10 core genes were missing in the analysis;**
- **Above 24k annotated protein-coding genes;**
- **Our study is the first large-scale report on the Ricciaceae nuclear genome.**

Funding: This work was supported by The National Science Center Kraków, Poland: Grant No. 2020/39/B/NZ8/02504