



INSTITUTE OF EXPERIMENTAL BOTANY
Centre of Plant Structural and Functional Genomics

Eva Hřibová, Ph.D.
Institute of Experimental Botany AS CR
Centre of Plant Structural and Functional Genomics
Šlechtitelů 31
779 00 Olomouc
Tel +420 585 238 703
hribova@ueb.cas.cz

Review of the Ph.D. thesis of Tihana Vondrak entitled

'Analysis of repeat-rich regions of plant genomes using long read sequencing technologies'

Introduction part of the thesis is rather short, but gives a brief overview of the satellite DNA, its origin and evolution and methods which have been used in the satellite DNA research. The results which follows the Objective and aims of the thesis, are presented as three publications in the high impact peer-reviewed journals, with Tihana Vondrak being the first author for two of them.

First paper describes the development of a work-flow for detection and genome-wide analysis of the satellite DNA arrays in unassembled nanopore reads in the grass pea (*Lathyrus sativus* L.), a plant species with a relatively large genome and a small number ($n=7$) of meta-polycentric chromosomes and 23 putative satellite DNA families. Long nanopore reads enabled to investigate the patterns of the array size distribution, their sequence homogenization pattern, and possible association between the different types of satellite DNA families. Besides other things, it was showed that the most investigated DNA satellites occurred in the genome as short tandem arrays (< 5kb), and originated from the repeats present in 3' untranslated regions of the Ogre retrotransposons. Preferential expansion of retrotransposon-derived satellites in the pericentromeric regions of *L. sativus* was showed and potential mechanisms of their pre-dominant location in primary constrictions were discussed.

Šlechtitelů 31, 779 00 Olomouc-Holice
Tel.: 585 238 703; Fax: 585 238 704
E-mail: hribova@ueb.cas.cz
<http://olomouc.ueb.cas.cz/>



In the second paper, nanopore reads longer than 30kb (N50 ~ 57 kb) and representing 5 x coverage of the genome of holocentric plant *Cuscuta europaea* were used to investigate complex structure of heterochromatic regions containing major satellite DNA repeat CUS-TR24. Thanks to the investigation of the long nanopore reads, a complex repeat structure CUS-TR24 array regularly interspersed with SSR repeat and the targeted insertion of LINE transposable element was revealed.

The last aim of the thesis was achieved in the third publication, published recently in the Cell journal, where Tihana is a co-author and focuses on the evolutionary dynamics of centromeric repeats of holocentric species *Rhynchospora pubera*. With this aim, chromosome-scale assembly of *R. pubera* and two other holocentric (*R. breviuscula* and *R. tenuis*) and closely related monocentric species (*Juncus effusus*) were analyzed. Autooctoploidy of *R. pubera* enabled her to study polymorphism of centromeric satellite Tyba in four paralogous copies in the genome, and clear link between CENH3 deposition and Tyba satellite was proposed.

The Ph.D. thesis of Tihana Vondrak is of excellent quality and brings significant advance in the study of the long-range organization of satellite DNA, its origin and evolution by development and application of a new bioinformatic pipeline for the annotation of repeats in the long unassembled nanopore reads. Thus, I'm glad to recommend the thesis for the successful defense.

Questions to be discussed during the Ph.D. defense:

Long nanopore reads containing satellite DNA were identified based on the homology search against an already existing database of DNA repeats, usually provided in previous studies by de novo repeat identification from the low coverage illumina data using RepeatExplorer.

Is there any other workflow for de novo identification of satellite DNA (tandemly organized) sequences in the long reads data (assembled or un-



INSTITUTE OF EXPERIMENTAL BOTANY
Centre of Plant Structural and Functional Genomics

assembled) without a need of homology search against the previously characterized repeats? If so, did you try to use it for the same data you analyzed during your PhD study and can you compare the advantages and disadvantages of such approaches?

In your work, error-prone nanopore sequences were analyzed without any assembly or correction step to find out the structure of large satellite arrays. Is it possible to analyze also sequence diversity within the satellite arrays if nanopore reads are corrected (by assembly process or by illumina data)? What is the general error rate of raw nanopore data which were used in your study, compared to PacBio HiFi reads?

Olomouc, August 30, 2022.

Eva Hřibová, Ph.D.

INSTITUTE OF EXPERIMENTAL BOTANY
Centre of Plant Structural and Functional Genomics
Šlechtitelů 31, CZ 779 00 Olomouc
Czech Republic
VAT No. CZ61389030, tel.: +420 585 238 702

Šlechtitelů 31, 779 00 Olomouc-Holice
Tel.: 585 238 703; Fax: 585 238 704
E-mail: hribova@ueb.cas.cz
<http://olomouc.ueb.cas.cz/>

Univ.-Prof. Dr. Hanna Weiss-Schneeweiss
RG Plant Evolutionary Cytogenetics
Department of Botany and Biodiversity Research
University of Vienna
Rennweg 14, 1030 Vienna, Austria
<https://cytogenetics.univie.ac.at/>
Tel: +43 1 427754159
hanna.schneeweiss@univie.ac.at



Vienna, 05 September 2022

Report on the PhD thesis of MSc Tihana Vondrak

“Analysis of repeat-rich regions of plant genomes using long read sequencing technologies”

The thesis focuses on development and testing of novel bioinformatics approaches to analyze the tandemly repeated satellite DNAs in plant genomes. Using long sequence reads and/or genome assemblies based on long sequence reads the thesis aims to develop novel bioinformatics approaches to identify and perform in depth analysis of the organization of satellite DNAs in selected plant genomes/species, with emphasis on their origins and evolution, partly in relation to their function(s) in the genomes. The organization of satellite DNAs is analysed in species representing different phylogenetic lineages and differing in centromere organization and its impact of overall genome structure and evolution is tested.

The thesis is very well written, clear to read, and free of typographical and other errors. The thesis addresses three main topics: (1) development of the bioinformatics pipeline for identification and annotation of tandem repeat arrays using ultra-long Nanopore DNA sequence reads; (2) testing the previous hypotheses on patterns of long-range organization of selected satellite DNAs in two plant species: *Lathyrus sativus* (Fabaceae) and *Cuscuta europea* (Convolvulaceae,); and (3) comparative analyses of the effects of holocentricity (with emphasis on centromeric tandem repeats) on whole genome organization (compartmentalization) in three related holocentric *Rhynchospora* species and their monocentric relative *Juncus effusus* (Cyperaceae).

The thesis encompasses three papers, two published and one accepted. All papers are published/accepted in very good scientific journals. In two of the papers the PhD candidate is listed as first author, and in the third one her contribution is substantial.

Introduction summarizes the topic of the thesis very well, is factually correct, very well written and concise. It provides all information necessary to understand the aims of the study and the results presented in the thesis. Literature is up to date and cited correctly. The Introduction consists of three parts. First part introduces repetitive DNA types, with the emphasis on satellite DNAs, their origin, properties, mechanisms of evolution and organization in the genomes. Part two focusses on the types and organization of centromeres and part three provides succinct overview of the methods employed in satellite DNA research from the beginning up to most recently developed bioinformatics RepeatExplorer pipeline. The Introduction very logically leads to the aims of the thesis (summarized above).

The **first Chapter (paper published in *The Plant Journal*)** presents novel assembly-free approach to infer the organization of 11 major satellite repeats using ultra-long Oxford Nanopore reads in *Lathyrus sativus*, species with meta-polycentric chromosomes. Two general types of satellite DNAs organization have been identified among the selected 11 satellite DNAs, challenging the currently accepted views on tandem repeat array organization. Only two satellite DNAs (one centromeric and one subtelomeric) were found to be organized in long tandem arrays, whereas the remaining nine satDNAs (all localized within primary constrictions of the chromosomes), were inferred to be derived from short repeats within Ogre LTR-retrotransposons and organized in short(er) or intermediate-sized arrays. The in-depth analyses provided wealth of data and the data were interpreted and discussed in the light of relevant published data. Novel hypotheses were proposed.

Questions/comments:

1. The two contrasting types of organization of satellite DNAs have been discussed in relation to their age and location in the genome, as well as functionality. What is your opinion on the impact of these factors on the types of organization observed? Which of these (age of repeats, their genomic location and/or functionality, recombination frequency of the genomic regions they are located in) may in your opinion play more important role in creating such distinct organization types?
2. How may the clearly demonstrated inter-relation of transposable elements and (at least some) satellite DNAs impact the accuracy of annotations of satellite DNAs using Illumina reads?
3. One system in which organization of satellite DNA can also be analysed are B-chromosomes which not only can evolve novel satellite DNAs, but sometimes show different patterns of organization of satellite DNAs present also in A-chromosomes.

The **second Chapter (paper published in *Computational and Structural Biotechnology Journal*)** is a logical follow up of the first chapter and sets to examine the internal structure of the heterochromatic regions containing satellite DNA CUS-TR24 in the of holocentric *Cuscuta europea* chromosomes using the pipeline developed and published in paper 1. The results revealed that the arrays of this satellite DNA are interspersed with SSRs and fragments of mobile genetic elements (LINEs). This is a very elegant study that examines the in-depth organization of heterochromatic regions containing CUS-TR24 and interpreting the data in very broad context. The model of evolution of such arrangement is very informative and well supported by data.

Question:

1. Are centromeric regions of holocentric chromosomes also recombination cold-spots?

The **third Chapter (paper accepted for *Cell*)** presents first chromosome-scale assembly of holocentric genome of genus *Rhynchospora* that possess repeat-based centromeres. The assembly is then used for comparative analyses of organization and evolution of paralogous centromeric loci of three holocentric *Rhynchospora* species and their relative, monocentric *Juncus effuses*. The particular focus of this study is on the patterns and mechanisms of karyotype restructuring accompanying speciation in relation to centromere organization via chromosomal fusions (EEFs). The role of centromeric satellite DNA Tyba (and its relation to other type of mobile genetic element, Helitron) is also analysed in-depth. The study is carefully designed and conducted to highest standards. It clearly demonstrates the impact of centromere organization on the genome architecture.

The thesis of Ms Vondrak tackles old questions/hypotheses with novel approaches and provides new important insights into the nature, origin and evolution of satellite DNAs, at the same time testing some of the factors underlying these patterns. This PhD thesis certainly contributes significantly

towards a better understanding of the role of repeats in genome evolution in plants. At the same time, the results provide many new questions and hypotheses to be tested in future. Ms Vondrak, as the first author of two publications and significant contributor to the third paper, manifested her ability to conduct independent scientific work. Collectively the thesis presents a very substantial body of work, carried out to highest international standards. I truly enjoyed reading the thesis and all its chapters.

I attest that the submitted PhD thesis of Ms Tihana Vondrak fulfils all general standards of PhD dissertation and is a fitting basis for awarding Ms Vondrak the degree of "Doctor of Philosophy". I warmly recommend the thesis for the defense and I also congratulate Ms Vondrak on the excellent work.

Yours sincerely

Hanna Schneeweiss