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 Laura Ávila Robledillo entitled

'Evolutionary dynamics of satellite DNA in plant genomes'

The thesis of Laura Ávila Robledillo focuses on genome-wide analysis of satellite DNA in the genome of Fabeae species. She studied the origin, evolution and function of satellite DNA, their association with functional centromeres as well as the organization of different types of satDNA in the nuclear genome, on chromosomes. The methodology which was used during her studies is state-of-the art: she took advantage of a recently developed computational pipeleline (Repeat Explorer) with the combination of molecular and cytogenetic methods to provide comprehensive characterization of satellite DNA. Furthermore, she applied a long reads sequencing technology – Oxford Nanopore sequencing to analyze the organization of repetitive units within the satellite arrays, patterns of sequence homogenization and their association with other DNA sequences. To do this, she developed a workflow for the identification of satellite DNA arrays in long nanopore sequencing reads.

The thesis is clearly structured and concisely written. The Introduction part gives an overview on satellite DNA in plant species, their composition, origin and evolution with the focus on centromeric satellites, and briefly describes methods used for satellite DNA identification and characterization. The English language of the thesis is as good as in scientific papers.

INSTITUTE OF EXPERIMENTAL BOTANY
Centre of Plant Structural and Functional Genomics

Laura Ávila Robledillo brings new information on the complex pattern of arrangement of satellite DNA arrays in Vicia faba and other 13 Fabeae species. Her results have pointed out to independent origin or rapid sequence diversification of satellite DNA among the Fabeae species. By application of ChipSeq analysis, she has detected the species specific satellites, associated with centromeric chromatin, which differ among species or even among different chromosomes of the same species. Based on the results, she has proposed a scenario for the evolution of centromeric repeats in Fabeae tribe. Finally, she applied the long read sequencing on investigation of the long-range organization of satellite DNA of Lathyrus sativus. Besides other things, she showed that most of satellite DNA of L. sativus originated from the short tandem repeats in the 3' untranslated region of Ogre retrotransposons and were predominantly located in the primary constriction of metaphase chromosomes.
Key results of her analyses were published in three high impact research journals, with Laura Ávila Robledillo being the first author for two of them.

Without any doubt, the Ph.D. work of Laura Ávila Robledillo brings new important results on the organization and evolution of satellite DNA in plants, Fabaeae family in particular. Therefore, I approve on successful Ph.D. defense.

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

1. In the Introduction, the benefits of long read sequencing technologies for comprehensive studying of large-scale arrangement and sequence diversity of satellite DNA is mentioned. Are there any other approaches which can be used to analyze the large-scale diversity and organization of satellite DNA?

2. In all three publications, the combination of FISH with immunolabeling was used to find out the association of specific satellite DNA with the centromeric chromatin. Would it not be more appropriate for this purpose

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/

INSTITUTE OF EXPERIMENTAL BOTANY
Centre of Plant Structural and Functional Genomics

to apply a technique of super-resolution microscopy which can provide more detailed information about the chromatin structure?

3. In paper, 'Characterization of repeat arrays in ultar-long nanopore reads reveals a frequent origin of satellite DNA from retrotransposon-derived tandem repeats", a workflow for the satelitte DNA sequences identification in long nanopore reads was based on homology searches with the already known tandem repeats originally identified in Illumina data using Repeat Explorer pipeline. Have you tried also an approach which enables to identify tandem arrays in nanopore read (or in their contigs/scaffolds) de-novo? Do you expect that de-novo approach might lead to the identification of new putative satellites which were not discovered by Repeat Explorer pipeline in illumina data?

Olomouc, June 6, 2021.

Eva Hřibová, Ph.D.

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/

TECHNISCHE UNIVERSITÄT DRESDEN

Fakultät Biologie, Institut für Botanik, Professur für Zell- und Molekularbiologie der Pflanzen

Technische Universität Dresden, 01062 Dresden

Dr. rer. nat.
Tony Heitkam

Telefon: 0351 463-39593
Telefax: 0351 463-39590
E-Mail: Tony.Heitkam@tu-dresden.de

14.06.2021

Opponent's review of the PhD thesis of MSc. Laura Ávila Robledillo "Evolutionary dynamics of satellite DNA in plant genomes"

Topic and objectives: Ms. Ávila Robledillo hands in a comprehensive thesis of high quality, covering a timely topic in the area of plant genome biology and evolution. Currently, even in "gold standard" genome assemblies, satellite DNA (satDNA) is still represented insufficiently and erroneously. The thesis at hands explores how a combination of sequencing, bioinformatics/genomics, and cytogenetics can shed light on satDNA origin, evolution, and biological role, especially in the centromeric chromatin. For this, Ms. Ávila Robledillo focuses on a number of species in the Fabeae, especially on *Vicia faba*, to (i) comprehensively characterize the repeat landscape in *Vicia faba*, especially in satDNA diversity and genomic localization; (ii) to analyze the contribution of satDNA to the centromeric chromatin, especially in closely related species; and (iii) to investigate the origin of satDNA using long read technologies. The objectives are well defined and result directly from the topic and research questions.

Structure of the thesis: The submitted thesis is structured logically and the chapters are well-connected, building onto each other. The thesis is cumulative, starting with a combined introduction. The introduction is followed by the "Aims" and "Scope" sections, outlining the thesis aims and framing the individual chapters of the thesis. These chapters consist of three published research works that have appeared in the peer-reviewed, internationally renowned, and well-regarded journals "Scientific Reports", "Molecular Biology and Evolution", and "The Plant Journal". Ms. Ávila Robledillo is lead author for two of the published works, and second author for the third article.

Formal criteria: Ms. Ávila Robledillo always uses correct grammar and spelling, and always accurately employs the correct technical terms of the research field. All terms used are explained and the reader is able to follow the content without any problems. The literature is formatted correctly. All of this convincingly illustrates that Ms. Ávila Robledillo knows and fittingly applies the formal standards of her research field.

Introduction, aims, and scope: The dissertation starts with a combined introduction spanning eight pages and approx. 120 references, covering the most pressing research questions regarding satDNA biology, their evolution, and historical/current approaches towards their detection.

Ms. Ávila Robledillo begins with a chapter on the sequence composition and abundance of satDNAs in plants, and focuses on copy number fluctuations often noticed between related plants. With this, she already sets the scene for her research, highlighting the question how satDNAs may diverge between species, also at different timescales. She then addresses the possible origin of satDNAs. This part includes a very small inaccuracy, as the integration of satDNA into the rDNA is referred to as transposition. Ms. Ávila Robledillo then moves to several hypotheses of satDNA evolution and describes their impact on the research field, focusing especially on concerted evolution. In this context, the debated satellite library model could also be discussed and weighed against the other hypotheses/models. For the defense, I would also be interested to learn more about the possible effects of segmental duplications and gene conversion of satDNA evolution. Ms. Ávila Robledillo then moves to explain the contribution of satDNA to the centromeric chromatin, and outlines the most widespread assumptions surrounding this topic, such as the centromeric drive model.

Ms. Ávila Robledillo finishes the introduction with a summary of the techniques for satDNA identification and characterization, covering classical molecular biology techniques and newer genomics/sequencing approaches, including next-generation and long read sequencing.

Regarding satDNA, I consider the introduction as complete and very clear, including all necessary references to understand and appreciate Ms. Ávila Robledillo's research. I would have liked to also read an introduction into the biological model, *Vicia faba*, and the related Fabeae as well as the state of the art of repeat genomics/cytogenetics in these plant species. A general insight into the state-of-the-art of cytogenetics for repeat analysis may have been possible as well. Nevertheless, the introduction works very well as it is and provides a perfect background to understand the following chapters.

The aims are fitting, resulting directly from the research questions, and clear. The scope connects the three chapters and papers and provides additional explanation. A combined scope/discussion section, unifying the results of the three

papers and explaining the combined contribution to plant biology would have perfected the overall very good impression.

Methodology: Ms. Ávila Robledillo applies a wide range of techniques, always combining computational with cytogenetic techniques, and challenging the state-of-the-art, e.g. by providing a deeper satDNA analysis as is usually performed in Chapter 1, or by comparatively analyzing the centromeric contribution of satDNA in 14 species of Chapter 2, to my knowledge the most comprehensive comparative analysis of comparative CENH3 ChIP-seqs in plants. When long read technology became available, she did not hesitate to include this into her work as well (Chapter 3). All Chapters are performed in a collaborative, often international team and testify to the ability of Ms. Ávila Robledillo to perform in diverse working groups and to find expert help, if needed, e.g. by cooperation.

Findings and conclusions: The results presented are diverse, plentiful, and novel. Bioinformatics analyses are always complemented by *in situ* hybridizations of highest quality, thus satisfying both, the algorithmic identification and the experimental validation aspect. If applicable, figures and tables are included to illustrate the most important findings. These are clear in interpretation, correctly labelled, and correspond to the highest standard.

Ms. Ávila Robledillo has been able to successfully group the findings to produce three timely and very interesting publications. These have passed peer review and are of excellent quality, as also indicated by the high reputation of the respective journals.

Depending on the time given for the defence, the many, high-quality results provide an optimal starting point for a deep discussion in the defence, e.g. focusing on suspected reasons for the variable satDNA monomer lengths, the clustering of different satDNA families, and the observed differences in replication timing (Chapter 1); or on the different contributions to centromeric chromatin, the absence of detectable centromeric repeats in three species, and how this situation compares to other species such as the grasses (Chapter 2). Chapter 3 gives also many interesting insights: The newly developed methods to investigate long reads also opens the way for interesting discussions, e.g. on possible biases that go back to the sequencing procedure, the interpretation of some of the graphs (e.g. the lack of symmetry of Figure 3 or how to interpret the spectra of Figure 5), and the general implications for the origin of satDNA, especially weighing the different hypothesis for satDNA evolution.

Further possible questions for the defence of Ms. Ávila Robledillo: Next to the possible questions that I have addressed in the "results" paragraph, I am outlining

further possible questions that have arisen when consulting the thesis and that may contribute to the defence of the thesis:

(1) Please outline why *Vicia faba* was chosen as a reference organism in the first study, what the state of the art on *Vicia* repeat genomics was before embarking on the thesis, and how this thesis has contributed to advance knowledge of Fabeae genomes.
(2) Please put into a wider context how comparable to other plant genomes the findings are regarding satellite diversity (fist manuscript), their contribution to centromeric chromatin (second manuscript), and their possible origin from other genomic components, e.g. retrotransposons (third manuscript).
(3) Regarding concerted evolution and the satellite library model, can these hypotheses be explained and weighed against each other? Are there possibilities to recognize how segmental duplications and gene conversion affects satDNA evolution?
(4) For Chapter 1, why are frequencies of di- and trinucleotides in the satDNA consensuses measured?
(5) As mentioned above, for Chapter 3, I have some unresolved questions regarding the interpretation of Figures 3 and 5 that may be discussed at the defence.

Overall assessment: The presented work of MSc. Laura Ávila Robledillo is one of the most comprehensive reports of satellite DNAs in related plants, especially regarding their contribution to the centromeric chromatin. It covers satDNA identification and characterization, their localization along the chromosomes, and answers complex evolutionary biology questions regarding satDNA origin and role in centromere formation. All of the thesis' objectives have been addressed and the thesis aims were fulfilled at a very high scientific level. The thesis convincingly demonstrates Ms. Ávila Robledillo's ability to work scientifically to highest standards, yielding many new insights that she has also turned into three strong publications. The small points that I have indicated and may provide excellent discussion topics in the defence and I want to make clear again that these do not diminish this thesis in any way. As such, Ms Ávila Robledillo's thesis clearly fulfils all criteria needed and I fully and without any hesitation recommend it strongly for the defence procedure.

Tony Heitkam

Dr. Tony Heitkam
Dresden, 14.06.2021