Review of the Ph.D. Thesis

Mgr. Jaroslav Zahradníček: Phylogeny and evolutionary trends in subtribe *Hieraciinae* (*Asteraceae*)

The thesis consists of brief general introduction to the studied group, three published papers (all in quality scientific journals with IF), one manuscript intended for publication in a journal with IF, and finally very brief conclusion. The candidate is the first author of one of the published paper and the manuscript. Two papers focus on genome size variation of the studied genera (*Hieracium* s.str. and *Andryala*) and its correlation with phylogeny and selected life history traits. These two papers have similar structure, methodology and analytical tools used. One paper is focussed on phylogeny of the little known and intricate genus *Andryala*. The last paper reviews ploidy level variation, breeding system and genetic variation (phylogeography) of one peculiar species, *Hieracium intybaceum*. The four studies well fit together thematically. All of them bring new primary data (genome size estimates, DNA sequences, etc.) that are analysed in standard ways and provide original results that significantly improve our knowledge of the studied group.

In the light of these information, it is clear that the candidate proved his ability of scientific work and knowledge of the methods used and the thesis fulfils current standards for Ph.D. theses in systematic botany. Nevertheless, I must also note that the chapters Introduction and Conclusions are brief and should be more focussed on the thesis topic. In Introduction, while the part on the studied plants is detailed enough, parts on polyploidy, apomixis and genome size are rather general and details specific to the studied group are scarce. The short Conclusions chapter is a summary of the individual papers rather than a real conclusion linking individual pieces of information.

Questions for discussion during the defence are listed below. Most of them relate to Paper III, in which the candidate is the first author, and Paper IV (manuscript), that has not undergone the review process yet.

1) Generic classification of *Hieracium intybaceum* / *Schlagintweitia intybacea*. There are several reasons for separating from *Hieracium* (molecular data, morphology) as well as for retaining in *Hieracium* (hybridization) but there also are objections against all of these reasons (incongruence of different markers; chloroplast capture events, etc.). What is the candidate's opinion (and reasoning)?

2) In both studies focussed on genome size, significant variation is found within several species both in *Hieracium* s.str. and *Andryala*. While there is detailed discussion on this topic in *Hieracium* (Paper I), there is no additional information in *Andryala* (Paper IV). Is this variation confirmed by presence of bifurcated peaks in simultaneous analyses of putatively divergent individuals? What may cause such variation (in the observed cases; I am not asking about various theoretically possible mechanisms)?

3) Genome size evolution. The genome size evolution was analysed in *Hieracium* (Paper I) and *Andryala* (Paper IV) in very similar manner using BayesTraits software. However, there is one important difference – in *Hieracium* collection of ~1500 individual trees was used (to account for phylogenetic uncertainty), while in *Anrdyala* only one consensus tree (= no uncertainty) was used. Why the different approach was chosen? Many branches in the *Andryala* tree have week no support. To what degree might this influence the results? Similarly, in *Andryala* differences between continental and insular and between short-lived and perennial taxa are tested. This would have real meaning only if phylogeny was taken into account (using methods like phylogenetically independent contrasts). If the species in a group (e.g. annuals) are related (forming a clade), the difference can be just inherited from the common ancestor and more or less random – no functional explanation is possible. Might this pattern (relatedness within a group) be present in the current dataset?

4) Two papers are focussed on the genome size variation within the genera *Hieracium* and *Andryala*, respectively, plus there is a similar older paper on *Pilosella* (Suda et al. 2007). Are there some similarities

among the three *Hieraciinae* genera, is some generalization possible (for example 'basic' vs. hybridogenous species, sexual vs. apomictic taxa, overall inter- and intraspecific variation, relation to life forms, etc.)?

5) In the thesis Introduction and in introductions of Papers I and IV, various functional (physiology, growth rate, cell metabolism, etc.) explanations of genome size variation in plants are emphasized. To which of these factors / processes the (part of) the observed variation can be attributed, in the candidate's opinion?

Minor & technical comments and suggestions:

– Introduction, p. 10. The genus having the smallest genome size is *Genlisea*, not *Gentiana*.
– Paper I and IV, Table I (surprisingly also in published paper). Variation (SD, SE, min, max) is reported also for taxa/populations that have only one accession – either by mistake or showing zero (for standard deviation etc.) but this is nonsense, there should not be any value in such cases. Revise the whole table in Paper IV before submission.
– Paper III. It would be useful to see the results of Structure analysis for K = 2. It is written in the text that it is one of the two possible solutions, but the results are not even mentioned – it should be at least shorty discussed what belong to each groups, if it makes some sense, etc.
– Paper IV. The unusual population *A. ragusina* II is highly interesting. When compared to 'normal' *A. ragusina* I, the genome size well corresponds to a putative triploid. Although it would be the first polyploid (and perhaps apomict) in the genus, such possibility cannot be completely ruled out, especially when such species are known from related genera.
– Paper IV. All analyses (continental vs. insular species, etc.) of the complete data set that involves the two divergent basal ('relict') lineages are useless as the significant results are mainly caused by these outliers (the difference between these two and the 'major radiation group' is much bigger than differences within the 'major radiation group'; from purely statistical point of view, these two species hugely violate homogeneity of variance assumption of ANOVA and t-test).
– Paper IV. Figure 2 is especially ugly: the data covers less than 50% of the y-axis (= most of the graph is empty space) and any differences in the genome size in the main group cannot be visible; moreover, resolution (dpi) of the printed version is insufficient. In my opinion, Fig. 3 with species ordered by their genome size (instead of alphabetically) will show the same pattern more clearly and with additional information on intra-specific variation.
– Paper IV (and also the others). When using box-and-whisker plots, it is necessary to indicate (in the plot or in the figure caption) what values are actually drawn: median or mean, quartiles or standard deviation or standard error, min-max or some confidence interval.
– Paper IV (and also the others). Sheffé's test is generally not recommended as a method of multiple comparisons (after one-way ANOVA), especially in cases when many means are compared. There is no optimal solution but Tukey HSD test or some other methods are more reliable.
– Paper IV and Paper I. Transformation of the genome size data. From methods, it seems that the data were $\log_{10}$ transformed for analysis using BayesTraits. For ANOVA and t-tests, it is not clear but it seems that there was no transformation for ANOVA (Paper IV, p. 4) but the data were transformed for t-tests (p. 5). As both methods have the same assumption, the data should be treated in the same way; log transformation cannot be used mechanically and the character of the data must be studied first (log-normal distribution, dependence of variation on mean, etc.); if the data does not require it, the transformation has negative impact on the analysis.
– Paper I. The use of linear correlations in such cases should also be re-considered. The trend might be non-linear (for example logarithmic or even unimodal (e.g. quadratic) with some optimum and decrease to both sides) or some transformation (such as log) of one or both variables might be needed to make the trend linear. All these models are also sensitive to outlying points (that have high 'leverage'), such as the two easternmost localities at Figs. 6 and 7.

České Budějovice, 20. 8. 2017

Mgr. Petr Koutecký, Ph.D.