BOTANICKÝ ÚSTAV SLOVENSKEJ AKADÉMIE VIED

Mgr. Judita Zozomová, PhD.
Botanický ústav SAV, Dúbravská cesta 9, SK-845 23 Bratislava
Tel: +421 904 434 770, E-mail: judita.zozomova@savba.sk

Bratislava, 2 July 2012

Review of the PhD thesis of Karol Krak: Molecular phylogeny and evolutionary trends in *Hieracium* (Asteraceae, Lactuceae)

The presented thesis is devoted to the phylogenetic and evolutionary studies of the subgenus *Hieracium*, which is a complicated agamic complex with large morphological variation, extensive past hybridization, and contradictory taxonomic concepts. Previous studies concentrated mainly on describing species diversity and reproduction, whereas species relationships have not yet been comprehensively addressed and remained unexplored. Therefore, the aims of the thesis – to infer phylogeny, delimit major evolutionary lineages, disentangle hybrid and polyploid origins – represent highly relevant and challenging tasks. The approach adopted here, i.e., to include all Zahn's 'basic' species and employ multiple molecular markers characterized by different inheritance and molecular evolution, was sound, promising comprehensive and convincing phylogenetic inferences.

The thesis is composed of introductory chapters and four scientific papers (in which K. Krak was either a principal author or a co-author), three of them already published (in renowned journals), and one in the submission stage. **The introductory chapters** are devoted to 1) evolutionary significant processes and features that have played major roles in the *Hieracium* evolution: hybridization, polyploidization, and apomixis; 2) review of molecular markers, with the focus on those employed in the thesis, and highlighting their strengths but also drawbacks and potential pitfalls; 3) overview of the studied (sub)genus, considering its reproduction, taxonomic aspects, and reviewing so far very scarce phylogenetic studies. All these chapters are concise and very well focused; I appreciate especially the section on molecular markers that is presented comprehensively, stressing really the most relevant and interesting points, and showing a remarkable grasp of the subject by the author.

Paper 1 (BMC Evolutionary Biology) concentrated on resolving phylogenetic relationships and inferring hybrid and polyploid origins in *Hieracium* s. str. based on ribosomal ETS and cpDNA sequence data. The data analysis was definitely not straightforward, due to the high level of intra-individual ETS polymorphism and lot of misleading variation. Probably many other researchers would fail and completely discard such a dataset. But the authors apparently took this as a challenging task, and succeeded in discerning informative variation (i.e., the phylogenetic signal) from noise, stochasticity and homoplasy using a thorough inspection of the observed patterns and detailed character state analysis. In stark contrast to the expectations, the authors discovered extensive reticulation in the past, that concerned also basic, supposedly non-hybridogenous diploid species, and inferred a large diversity of ancestral diploids that gone extinct but left traces in the present taxa. **The paper brings completely new insights into the evolution of the subgenus.**

Paper 2 (Annals of Botany) analyzed nuclear genome size variation in *Hieracium* s.str. in a phylogenetic context. The distribution of genome size variation was congruent with the two major

ETS phylogenetic clades, suggesting that the evolutionary history is the major factor affecting and explaining the differences in genome size. Furthermore, it was shown that genome size can be helpful in species delimitations and circumscription; increased intraspecific genome size variation could in fact indicate inappropriate, polyphyletic species circumscription (in line with ETS heterogeneity) and/or polytopic origins of the polyploids and hybrids.

Paper 3 (American Journal of Botany) presents development of low-copy nuclear markers applicable for phylogeny reconstruction at low taxonomic levels within the subtribe Hieraciinae, having high potential also in other Asteraceae tribes. It is based on a sophisticated bioinformatic procedure utilizing sequence databases. The paper outlines the phylogenetic applicability of three developed markers, and certainly will **stimulate and facilitate future phylogenetic studies** in Asteraceae.

Paper 4 (Heredity) aimed at obtaining more phylogenetic resolution and deeper insights into the evolution of *Hieracium* s.str. by using two more variable markers – an additional intergenic spacer of cpDNA and a single-copy nuclear gene. The resolved clades and their relationships were quite incongruent among the markers, and this was attributed to either incomplete lineage sorting, or to extensive reticulations, although in some cases it was not possible to distinguish between these scenaria. The hybrid origin of many taxa as suggested by ETS data was confirmed, or even refined in several cases. The authors emphasized **the necessity of gathering data from multiple independent markers (and increasing intraspecific sampling) to infer species tree,** especially when recently diverged species groups with extensive reticulation are in focus**.**

All four papers are very well elaborated and conclusive (three of them in fact already passed the strict review process in the journals), so I mention here only a few comments or questions.

(1) The high number of analyzed species hampered deeper intraspecific sampling. Nevertheless, based on the patterns observed for the species where multiple accessions were examined (e. g. the results of *H. prenanthoides* in ETS), can you assess what might be the impact of increased intraspecific sampling, how this could affect the phylogenetic conclusions?

 (2) In the Results section of paper 4 I slightly miss some summary of parameter values of the datasets, such as the number of polymorphic positions, parsimony informative sites, number of indels inferred, number of haplotypes/alleles. This usually gives a general overview on the data used and might, although roughly, indicate also their phylogenetic potential. Regarding the squalene synthase sequences, I wonder, how different were the levels of polymorphisms between the exons and introns? Was the alignment of introns straightforward across the whole dataset? Have you tried to conduct a phylogenetic analysis based on exon sequences only? Have you inspected also the ratio of synonymous and nonsynonymous substitutions among different sequences (as the increased level of the second type can indicate the presence of paralogues)? Can you compare the phylogenetic signal/utility of exons and introns also for the other two markers (*shmt, gsh1,* paper 3)? For the detailed analyses in *Hieracium* (paper 4) you selected one of the three developed markers, the squalene synthase, can you elucidate what was your choice based on?

(3) Regarding cpDNA data (paper 4), you reported similar topologies of the consensus tree of the maximum parsimony analysis (a strict consensus or 50% majority rule consensus tree?) and the majority rule consensus tree of the Bayesian analysis, and referred to Fig. 1. Fig. 1, however, depicts one of the most parsimonious trees (MPTs) which is very likely more resolved than either of the consensus trees. Personally I do not think that presenting one of MPTs is the right choice, as it shows only one of the possible topologies and makes an impression of better resolved

relationships than it is true. I assume the clades B, C and D probably collapsed into polytomy in the consensus trees. For the *sqs* data you presented the Bayesian consensus tree, which I think should be followed also for the cpDNA dataset.

(4) Among the cloned sequences (papers 1, 4) you identified some recombinant sequences by visual inspection. What do you think about the efficiency of statistical approaches and algorithms implemented in some software applications (e.g., PhiPack) to identify such artificial recombinations? Have you ever tried them? Do you think, are they applicable also to complex datasets (with high levels of polymorphisms); could they outperform visual inspection?

   In conclusion, the presented thesis has brought substantially new insights into the speciation and evolutionary history of *Hieracium* s. str. The significance of these studies extends also beyond the study genus – they illustrate approaches that can be employed with success also in other agamic and polyploid complexes, and bring new markers applicable also in other related plant groups. At this point it is my pleasure **to recommend Karol Krak for the award of the scientific-academic degree "philosophiae doctor"**.

Judita Zozomová