



MINISTERIO
DE ECONOMÍA
Y COMPETITIVIDAD



CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



REAL JARDÍN BOTÁNICO (RJB)

Review of Karol Krak's PhD Dissertation: Molecular Phylogeny and evolutionary trends in *Hieracium* (Asteraceae, Lactuceae)

This Dissertation aims at throwing light on the origin of the diversity and relationships within a plant group with peculiar biological characteristics and mode of evolution. For this, it focuses on molecular markers and genome size, and specifically explores and analyzes nuclear markers (ribosomal DNA, namely ETS; and low copy nuclear genes) as well as plastid markers. This is presented along four papers, three of them already published, plus a good introduction summarizing the different relevant elements to study this system.

To appropriately frame and contextualize this review, it is necessary to stress that *Hieracium* is one of those groups of organisms that exemplifies taxonomic complexity and defies standard classifications. In fact, the British/Scandinavian school of hieraciologists consider that it cannot be ascertained how taxonomic entities in *Hieracium* were originated. Exploring the origin of *Hieracium* diversity does not solve the taxonomic problem but it may inform about particularly unsuitable taxonomic treatments and help understanding the diversity of modes and mechanisms of evolution in plants. To cope with complexity, different approaches and sources of evidence should be tried. But still the ultimate task of integrating the results from those sources is one of the toughest. Along this line, it is clear that a multidisciplinary approach is needed and thus the research work of the PhD candidate seems to be nicely complemented by a team behind. For instance, a thorough knowledge of the morphological diversity and how it is arranged geographically and taxonomically is essential for interpreting meaningfully the results from partial sources of evidence such as DNA sequence data or genome size. In this Dissertation, the analysis and interpretation on the history and relationships of every species sampled, based on molecular markers and genome size, have clearly benefited from a sound taxonomic knowledge of the genus.

I would highlight from this Dissertation the thoughtful and thorough effort to critically analyze and interpret the results. Because of the difficult system (*Hieracium*) and the specificities of the evolution of the genes analyzed, this task has required a detailed analytical strategy to scrutinize, clean and handle the data in order to extract signal from noise. The Dissertation contains relevant results that unveil unexpected findings (see below) and open new avenues for research. The four papers that integrate the dissertation are well written and the arguments defended are sound, the latter paper requiring still an ultimate round of editorial tidying up.

To summarize, in my opinion this PhD Dissertation gathers more than enough merits, has provided the candidate with a very good research training, fulfils the international criteria for obtaining the PhD degree, and certainly I consider that it is suitable for the defense.

I would like to highlight a number of relevant findings, among others:

1) The discovery of two lineages in the ETS phylogeny that seem to mirror the species phylogeny and were previously undetected.

2) A genome size pattern that matches those two ETS lineages. At first sight, one may think that genome size is a too dynamic or labile parameter to be useful in phylogenetic relationships among species. However, two arguments support the informativeness of genome size in this context. First, the fact that it is specifically the monoploid (Cx -) values what correlates with the ETS lineages. Secondly, the constrained possibilities for structural genetic changes in nearly-obligate apomicts as compared to sexual plants may provide the necessary level of stability to that parameter so as to be useful as a marker.

3) Previously undetected hybrid origins unveiled for a number of 'basic species'.

4) Reduced levels of concerted evolution in nr DNA ETS sequences, which would not be unexpected in organisms lacking sexual reproduction but requires explanation when it applies also to sexual diploids, as is the case.

5) The Dissertation provides information on species relationships, albeit generally not supported by more than one marker. Despite the somehow expressed frustration (particularly in paper #4), for the incongruence detected between cpDNA, nrDNA ETS and the low copy nuclear gene (LCNG) *SQS* sequences, in my opinion, it is not that dramatic when one considers the specific characteristics of each marker. When dealing with closely related species or (as in this case) a recently evolved genus it is not infrequent that the origin of some alleles of the gene predate the speciation events in question. This may be true even if the targeted sequence includes intron portions. If such is the case, one cannot expect a perfect match between genealogy of the gene and species phylogeny even if you are confident to be working with orthologs. This may help explain discrepancies between *SQS* and ETS lineages. As to plastid DNA, it has been amply realized (the Dissertation itself acknowledges this) that within closely related species, the patterns of variation often match geography rather than taxonomy (or species boundaries). Adding new plastid (thus linked) regions only reinforces the pattern. In sum, I do agree with the advocated hypothesis that ETS regions may convey the strongest signal on species phylogeny in *Hieracium* (called 'back-bone' in the Dissertation) as compared to *SQS* and plastid DNA phylogenies. It is not uncommon in other species-level plant studies to find that ribosomal DNA matches morphological (and thus classical taxonomic) schemes better than plastid or LCNG sequences.

On the other hand, I would downgrade strong statements about proofs of extinction of diploid species and ribotypes (e.g., in paper #1, but also elsewhere). The Dissertation makes a good point by indicating that, if apomicts are nearly obligate, their patterns of DNA sequence variation should be strongly based on their progenitors. But sampling, even if well designed, is after all limited both at the inter- and intraspecific levels. So, the possibility of unsampled genotypes cannot be discarded.

I have a minor discrepancy in some statements in paper #4 as to the processes that are responsible for lack of congruence among different markers. The PhD candidate and co-authors indicate that a comprehensive sampling (e.g. pag. 80) would be needed to tell apart introgression from patterns related to deep coalescence. This is true. But they go on to 'slightly' favor the hypothesis of reticulation over deep coalescence and incomplete lineage sorting to explain the *SQS* pattern (p. 81). I am not sure of this. Certainly, this is a difficult topic but reticulation is more feasible to be proved than lineage sorting. Along the discussion of paper #4, they single out cases in which reticulation has been involved. There are even some cases in which the *SQS* lineages provide some support for specific hypothesis of hybrid origin put forward based on ETS and cpDNA. So, reticulation is a likely cause for some of the patterns found as he rightly explains. However, in those other cases in which reticulation does not seem to be the determinant factor, such as the non-appearance of the western and eastern ETS

clades in the *SQS* tree (page 81), deep coalescence and incomplete lineage sorting is more likely to be the main cause behind, as explained above.

Finally, I have some questions that may well go beyond the scope of the Dissertation (some clearly) but that follow from reading it and I dare to write down:

- Why are recent hybrids so scarce (e.g., pages 12, 38)? Can you be so conclusive about this statement in cases with insufficient data, e.g., lacking chromosome number, sequences, but with pollen viability estimates available?
- What are the implications of the finding that some of the considered 'basic species' of the *Hieracium* taxonomic system may have a hybrid origin as suggested by the ETS and cpDNA data? Could the reverse be true, that some of the species considered intermediate (not sampled in this study and perhaps not thoroughly studied) may have contributed substantially to generate diversity and thus be considered 'basic'?
- What are the most determinant factors for the prevalence (no. of taxa) and ample distribution of apomicts as compared to sexual species (so called "geographical parthenogenesis")?
- What are the mechanisms underlying the Mentor effect and the genetic, taxonomic and ecologic consequences?

Gonzalo Nieto Feliner
Real Jardín Botánico, CSIC
Madrid, 2 August 2012