

Diana Libeth Aparicio Vásquez PhD – Evolutionary processes underlying plant diversification in the tropical Andean highlands

External Examiners Report – Colin E. Hughes

General assessment of the thesis

This thesis examines a set of research questions related to the evolutionary diversification of plants in the species-rich high elevation flora of the tropical Andes. The central aim of the research is to gain insights into the evolutionary processes underlying the remarkable concentration of recent and rapid plant radiations found in the high elevation Andean grasslands. In recent years many such radiations have been documented, but our understanding of how these radiations actually happened remains very limited. Of particular interest is understanding the role that Pleistocene glacial cycles with their associated cycles of flickering high elevation grassland habitat connectivity played in promoting rapid species diversification. These are the interesting and important questions addressed in this thesis. The research presented here provides important new insights into these questions and makes a significant contribution to our understanding of the two study groups, the genera *Lupinus* and *Senecio*.

These are difficult and challenging questions for several reasons. First, field collecting in the high elevation Andes is not easy due to the remote and isolated locations, the high elevations and the sheer geographical extent of the Andes which span several 1000km. Getting to know and collecting material of high elevation Andean plant groups is time consuming and demands resilience and fortitude. Second, the recency and rapidity of species diversification in the high elevation Andean grassland system poses particular challenges for obtaining robustly supported relationships using traditional molecular markers. Third, for many high elevation plant groups the taxonomy remains incomplete and chaotic; there are no recent taxonomic revisions for many groups; *Lupinus* one of the two study groups in this thesis, provides a prime example, and the generic delimitation problems associated with *Senecio* are another that the candidate has had to grapple with. This means that students studying these plant groups have to get to grips with the taxonomy themselves and grapple with issues of species identities. It is clear that the candidate has risen to these challenges, successfully carrying out field work in Ecuador and Colombia to assemble dense field sampling of a set of species, generating useful data and producing interesting and compelling results that contribute new insights into the processes underlying diversification in this challenging system.

This research encompassed a combination of field, laboratory and analytical / computational work which has involved learning a diverse set of relevant research skills and techniques in revisionary taxonomy, population genetics, demographic analysis and phylogenetics, including generation and analysis of three complementary types of molecular data: AFLPs, microsatellites and ITS sequences.

The emerging results of strong genetic divergence among populations, very low levels of within-population genetic diversity including heterozygote deficiency and isolation by colonization involving founder effects across the study species present a compelling hypothesis of high elevation Andean plant diversification.

Each chapter is accompanied by an extensive bibliography indicating generally very good coverage of the relevant literature. The candidate has clearly read widely around her research questions and study systems. Having said that, I found some notable and surprising omissions (see below).

The thesis is well written in good English and very few editorial mistakes were noted.

Chapter 1

This is a straightforward population genetics study to estimate levels of within and among population genetic diversity for *L. alopecuroides*. This species is well chosen as an excellent example of a high elevation species now occupying a highly disjunct sky-island distribution and the field sampling is thorough and includes all known populations of this interesting species. The data generated and the analyses are sound and the results are striking in terms of very low levels of genetic diversity within populations and high genetic divergence between populations, and the explanations (selfing, founder effects and bottlenecks) are interesting and make sense.

Research by Contreras et al (2018a,b) carried out subsequent to your 2016 paper suggested the COCUY population probably belongs in a different high elevation acaulescent rosette species, now named *L. luisanae* by Contreras et al (2018). It will be important to check back to the herbarium voucher material associated with the samples from Cocuy to check the identity of this material. It is striking that the only population in the 2016 study that shows higher genetic diversity is COCUY. Hence removing this population would render the results even more striking in terms of the patterns seen for *L. alopecuroides*. This mis-identification issue is mentioned in Chapter 2, but the COCUY population is still included in the analyses presented in Chapter 2, but should probably be removed in the final published version of this study.

It is not clear why the population from ANTISANA comes out as an outlier in the phenogram, suggesting it is genetically quite dissimilar from the other Ecuadorean populations and even more dissimilar than the three Colombian populations. The possible reasons for this are not properly explored or explained. What are the possible reasons for this? You mention founder effects can play a role in the initial stages of speciation and morphological differences associated with the ANTISANA material. Are you suggesting this could be a different species? An incipiently diverging species? Could there be gene flow with *L. nubigenus* which also occurs on ANTISANA?

Also, why was the ANTISANA population not included in the study presented in Chapter 2?

Chapter 2.

In this chapter the data from Chapter 1 are extended to include two additional species and a range of demographic modelling analyses are added. The results of this study present a compelling picture of the likely evolutionary processes underlying population divergence within species driven by post-glacial upslope isolation by colonization and is one of just a handful of studies of its kind for high elevation Andean plants. While these results are potentially significant and intriguing, there are a number of methodological issues that merit further attention and scrutiny before this chapter can be published. Specifically, the following issues need to be clarified:

- (i) if the species are autogamous, then surely this violates the assumption of free recombination among loci that underpins the coalescent simulation;
- (ii) is the default mutation rate in DIY ABC appropriate given that this default rate comes from rates measured in humans? Are there more appropriate mutation rates available that could be used? What about the estimates available from *Arabidopsis* and *Prunus* used by Nevado et al (2018) in their demographic analyses?

- (iii) you suggest that the priors used in the ABC modelling were wide and flat, but perhaps the number of generations (10 to 10,000) prior is too narrow and should span the whole of the last glacial period, i.e. 10 to 50,000?
- (iv) do you think the detailed correlation of split times with Holocene dates is justified given the uncertainties associated with these estimates? Was the LGM in the Andes not earlier than in the temperate zones?
- (v) no estimates of geneflow between populations were included in the ABC models that were tested – why not?
- (vi) The phenograms in Fig. 2 are suggested to depict ‘genetic relationships’ between populations of *L. alopecuroides* and *L. nubigenus*. Are these strictly speaking analyses of (phylo)genetic or phylogeographic relationships, or simply of genetic similarities? Your discussion of these phenograms implies a directionality (polarity) from ancestral to more evolutionarily derived populations. How do you infer that some populations are more derived than others? How are these trees rooted? Should they not be viewed instead as unrooted networks with potential ancestral populations falling on internal nodes rather than at the tips? Furthermore, given that the majority of the loci monomorphic within populations, it is not clear to me what are the shared derived character states that can be used to infer (phylo)genetic relationships? Without polymorphism how is it possible to infer evolutionary relationships? Furthermore, it is not clear how the SINCHOLAGUA population was inferred to be ancestral. It is not depicted as such.
- (vii) you used the phenograms depicting genetic relationships among species to formulate scenarios for the order of population divergences. Is divergence scenario 3 (i.e. phylogenetic tree) in Fig. 3 congruent with the phenogram in Fig. 2C? Similarly, for scenario 3 in Fig 4 – is that congruent with the branching order depicted in Fig. 2B?
- (viii) The recent landscape-scale habitat connectivity models developed by Flantua et al (2014) and Flantua & Hooghiemstra (2018) should be mentioned and used to help interpret the results. Similarly, the early paper by Simpson (1974) on paramo island biogeography is relevant.

Chapter 3.

This chapter focuses on species level diversification and phylogenetics of a different high elevation Andean plant group the *Lasiocephalus* / *Culcitium* clade of *Senecio*, using different molecular markers – AFLPs and ITS sequence data. It explores very interesting questions about the geographical and ecological trajectories of diversification, and the evolutionary shifts in growth forms frequently associated with high elevation Andean plant groups. Once again the sampling in this study is impressive and the results are interesting, albeit also tantalizingly frustrating due to lack of resolution in the resulting phylogenies and especially in the ITS gene tree.

The STRUCTURE analyses are elegant and informative, highlighting intriguing discordances between species limits and geography, indicative of the complex and potentially reticulating evolutionary histories of this group.

One wonders whether more cloning of ITS would have yielded interesting results, given the likelihood of divergent ITS copy types within individuals that would be expected, for recently diverged species, and especially for putative hybrids.

Chapter 4.

This chapter focuses on taxonomy and description of the new species *Senecio sangayensis*. Given how conspicuous these plants are, this is a startling and magnificent botanical discovery and it is excellent to see it competently described with adequate detail, excellent photographs, a distribution map and notes. Adding a botanical illustration would have been a worthwhile addition to illustrate details of the morphology that cannot be seen in the photos. It is not clear why no type material is recorded as being deposited in Ecuador? Overall this is a very nice small taxonomic contribution.

Conclusions

Overall this is a very good thesis, spanning a set of well-articulated and significant and important research questions, tackled using well sampled and pertinent study groups. There is a lot of new primary data and extensive analyses with interesting and novel results. In a few places, and especially in Chapter 2, further analytical work will be needed to bring this chapter up to the standard required for publication. With three papers published and one in revision, this thesis presents a substantial body of original research.

Questions

1. One of the questions that you focus on especially in your thesis is how Pleistocene glacial cycles influenced and potentially contributed to rapid species diversification in the high elevation Andes. Can you explain in detail what the possible hypothetical impacts of these cycles were likely to be in terms of evolutionary processes.
2. On what basis do you recognize the Páramo as distinct from the high elevation grasslands further south in the Andes? Is Páramo a meaningful and distinct biome for understanding evolutionary diversification of plants in the Andes as you suggest in your thesis? Are the plant groups that you have studied confined to the Páramo? What about other high elevation plant clades?
3. How do patterns of species diversification in the high elevation Andean grasslands compare with what we know about other tropical alpine mountains? Do we see evidence of similar rapid recent plant radiations? And if so, where? Do you think these other tropical alpine systems are likely to show similar patterns of flickering connectivity associated with Pleistocene glacial cycles?
4. You suggest that *L. alopecuroides* (and possibly also *L. nubigenus*) are semelparous and autogamous and highlight some possible reasons why you think this. Can you summarize the evidence for this. How convincing do you think the evidence is? How many flowers per inflorescence do these species have? Do you seed high fruit and seed set in these species? You hint at possible reasons why semelparous and autogamous species are favoured in high elevation Andean grassland habitats and indeed in tropical alpine habitats more generally. Why would this be the case?
5. Both *Senecio* and *Lupinus* have species that form giant stem rosettes. What other Andean genera show this growth form? It is also akin to similar growth forms in other tropical alpine habitats in e.g. Africa. What is the adaptive significance of this growth habit? And what features of the tropical alpine environment prompt these unusual growth forms?
6. What do you see as the most urgent and significant research priorities and questions for gaining further insights into the processes underlying explosive species diversification in the high elevation Andean grasslands? If you were given 0.5 million Euros (or even one million) today, how would you go about addressing these questions? What sort of data would you generate and why?
7. You talk about fragmentation and connectivity of populations of Páramo plants caused by Pleistocene glaciations in Chapters 1 & 2, but you don't attempt to explicitly quantify this in any way. What elevational shift was involved? How do we know this? How far back through the Pleistocene do we have data on this? What were the detailed quantitative landscape-scale impacts on habitat area and connectivity and have these been explored?

Related to this when was the last glacial maximum in the tropical Andes and did it coincide with the LGM in temperate zones?