DEPARTMENT OF ZOOLOGY
FACULTY OF SCIENCE
CHARLES UNIVERSITY IN PRAGUE

Viničná 7, CZ 128 44 Praha 2, Czech Republic

Review of the diploma thesis

Alžběta Poštulková
The role of the FAR5 gene in alpine adaptation of *Arabidopsis arenosa*

I have found the thesis exceptional. It comprises newly generated data and results that could easily serve as a base for several diploma theses. The combination of various complementary (but methodically very different) approaches is brilliant and provides reliable discoveries. I am amazed how far it goes. The candidate locus from a genomic scan is further dissected in detail using sophisticated clever crosses, experiments, and analytical tools. The design of labour-intensive crosses is splendid, and usage and understanding of advanced biochemical tools is admirable. I am convinced the thesis will serve as a base for a paper (or more papers) in a high-quality journal.

The thesis is logically structured. I was only surprised to find summary right after the introduction. Luckily it is summary of approaches and strategies and not the summary of the whole thesis. The text is concise, which is generally welcome. However, sometimes (mainly in the Methods part) it is so brief that it is hard to follow. The author supposes that readers are deeply familiar with all data, methods, and results of previous studies, which is not a polite approach to readers and the poor reviewer. For example, it is impossible to understand what "genomic dataset that includes genomes... from a total of 73 ... populations" means without a prior knowledge of the (six!) cited papers. It is not clear how many individuals were used, what proportion of genome was covered (perhaps not whole genomes, because PacBio assemblies were latter used for finding linkage of SNPs), how the data were generated... In few cases I was totally lost. For example, I had no idea what the five lineages were in the sentence "PicMin was applied on windows that had data for all five lineages". This number does not seem to fit anything from the previous parts of the Methods. The key for the above puzzle is found at the beginning of Results, too late for my taste. Perhaps the text is somehow prepared for a transformation into manuscript in which the Methods will be at the end, after the Results. However, this approach is hardly acceptable for the diploma thesis. Anyhow, I think that a small reshuffling of the Method and Result parts may help. Technical details like number of individuals fit better into Methods, I think. Details about PicMin (what it does and the size of window) are duplicated (they are both in the Result and Methods parts). The same holds true for the RDA. I am also not sure whether the meaning of A and F alleles should be explained only in the abstract. On the other hand, the statistical methods are nicely and precisely introduced and are easy to follow. Figures are carefully prepared and aptly chosen. They are easy to understand, and they fittingly illustrate the ideas.

I like the PCR-RFLP scoring, but according to my experience some reviewers may criticize it, because it lacks a positive control for restriction enzyme functionality, in other words, a longer amplified fragment or multiple PCR product, which is always cut. I also wonder how the manual scoring of band intensities in tetraploids was reliable. I think a gel photo illustrating the intensity differences would help.

I wonder what the *FAR5* particularly strong signal means. Was it stronger that in other loci? It is not possible to ascertain it from the Figure 1.

A comment about phylogenetic relationship of the *Arabidopsis* species (preferably a phylogenetic tree) would improve understanding of the ideas about ancestral states (pages 25-26). The ideas about changes in the whole Brassicaceae family without mentioning the phylogenetic relationships does not make any sense for me. Are the changes independent or just shared among closely related species? Perhaps the dataset can be used for finding molecular signatures of selection to better understand the evolution of the gene. Is it planned?

Why are the rare alleles called "non-native"? If I got it right the adaptations are orchestrated via increase of allele frequency from a standing variation. It means that the allele was there before. It is perfectly native! Maybe it is even selected for by some form of a frequency dependent selection or it was advantageous in the past. I wonder how the inferred important environmental factors fit to the situation in the past. For example, does it make sense to suppose that the A allele was favoured in lowlands during ice ages? Are the effective population sizes large enough to maintain polymorphism for thousands of years?

It seems to me that the discussion (and partly also introduction) is rather *Arabidopsis* centred. Perhaps it is OK for a paper manuscript, but generally we expect a larger scope in diploma theses. Does it mean that there are no other relevant papers about gene reuse, alpine adaptations, parallel evolution?

It is always great to cite some famous big names, for example the founders of modern evolutionary biology. It seems to me, that some parts of the thesis almost beg for such a citation. While the first paragraph of introduction reminds me a famous idea (theorem) of Fisher, the following text seems to fit with an idea (metaphor) of Gould. Does the thesis author anticipate what I am writing about? Anyhow, I appreciate that at least Kimura was cited.

I can conclude that the thesis fulfils all necessary requirements. The quality is outstanding. I am convinced it merits the highest possible evaluation.

Praha, the 16th of May 2024

Pavel Munclinger
Department of Zoology
Faculty of Natural Sciences
Charles University in Prague