Review of diploma thesis of Nikolas Tolar *'Genome assembly of Mechow's mole-rat (Fukomys mechowii) as a prerequisite for the analysis of transposon expression in mole-rat oocytes'*

I had the pleasure of meeting the author of the thesis during his presentations at the seminar for Bachelor and Master students, which were spectacular. Therefore, I was sincerely looking forward to reading this work. And I must confess that I was not disappointed. In fact, the thesis exceeded my expectations in several ways. I will mention these later in my review. I will not bother the committee or the audience with a description of the objectives, since they have already heard them during the presentation. Let me just state that I find the biology of naked mole rats and its overlap with human biology fascinating and I really like the topic. The work was definitely challenging (more like a smaller PhD project), but the prior interaction between Nikolas and his supervisor allowed for its successful completion. The thesis has a typical structure and is written in English. The quality of the language is excellent and surpasses the English of many PhD students I have dealt with. Perhaps it is too rich for a scientific text, the author sometimes does not use the simplest or most common phrases, but this is not at the expense of comprehensibility. I also appreciate the minimum of typos, awkward wordings and the authors' ability to clearly formulate scientific statements.

In the following text, I will discuss individual parts of thesis. The introduction is quite concise and covers both organismal and methodological parts of the thesis. It is a typical introduction of bioinformatic thesis, i.e. we do not see actual organisms, the only figure presented is an overview of the transposon classification. More figures would probably make the introduction more attractive, but given the quality of language and good text flow described above, it reads quite well. I also appreciate that there is a subchapter devoted to clearly stating the knowledge gap. On the other hand, I find the last part of the introduction describing the thesis structure redundant or at least coming too late.

Since much of the work is based on the use of fairly advanced methods, the corresponding chapter describing them is extensive and actually much longer than the introduction. I must admit that I am not an advocate of describing every component of the PCR reaction in minute detail in dissertations. However, since the ONT read assembly methodology described in presented thesis is still fairly new and developing, it makes perfect sense. I particularly like that in most cases the full version of the commands used are shown. Again, I would like to see more diagrams that would help a non-expert reader understand the logic of the data processing.

The results are divided into two parts. The first part deals with the description of Mechow's mole-rat genome assembly and annotation, the second with the analysis of transposable elements in selected mammalian species. In the first part, the author takes us on his successful journey to a satisfactory genome draft. I enjoyed the reading very much. In the second part, the flow is somehow lost, partly due to the complexity of the authors approach. It is still interesting, but requires focused reading with (at least in my case) frequent jumps back and forth. I felt like I was drowning in the multitude of graphs and tables, and understanding the differences between them took a lot of effort. It could be made more

concise and clear. Nevertheless, it is clear that Nikolas has mastered fairly complex bioinformatics approaches and, more importantly, that he can interpret complex data.

While reading, I jotted down my questions for the defense. Then came the Discussion and I was really frustrated because most of my questions were answered there. In my opinion, that is the best part of the thesis. At least to a certain extent. It definitely lacks a more detailed work with literature (only four cited references). Although far from being an expert on mammalian genomics, I doubt there is that little to work with. On the other hand, the author shows great insight into the context of his work. He is aware of possible shortcomings of the methods, discusses how to overcome them, and suggests future improvements and further experiments. Like most of the text, this part seems very mature and ready. I have little doubt that with a little more effort the thesis could be published in a decent scientific journal. Although I mentioned that most of my questions were answered in the Discussion, below are those that are still open. I would love to hear the answers during the defense.

Q1. I find the average ONT read length rather short. Have you checked the integrity of input DNA? Also, how would you improve the results of sequencing?

Q2. Is the resulting coverage sufficient? Author mentions recommended coverage 40x for de-novo genome sequencing. However, from my experience (although with repeat-rich genomes), often the 100x coverage and higher are necessary for chromosome-level assembly

Q3. Is it possible to further improve assembly with the existing data?

Q4. Were all the predicted transcripts/isoform considered for further analyses or was their (usual) redundancy filtered out? If so, how? If not, why?

From the above text, I hope it is clear that I enjoyed the thesis very much. Although I have some minor comments, I feel that it meets all the requirements for a thesis at USB. I always reserve the right to suggest the grade after the defence. But I can not resist to say at least: an excellent job!

České Budějovice, May 17th Aleš Horák