

Review of the master thesis of Eva Šochová
Evolution and genomics of symbionts in Hippoboscidae

The thesis of Eva Šochová is a relatively short text (cca 40 pages) organised in two research manuscripts ready to be submitted to a journals that follow after a brief general introduction of the theme. In both manuscripts Eva Šochová is the first author and her declared contribution is 70% and 60%, respectively. The first manuscript is focused on phylogenetic analyses of insect Hippoboscidae hosts and their bacterial endosymbionts based on several rRNA and protein markers. The second manuscript is centred on the genomic sequencing and annotation of two endosymbiont genomes. This is a first time I am reviewing a master thesis with this structure. I admit that the format is very practical, as it does not force the author to write long texts "only for the thesis", which will not be then published. On the other hand, I realize that this format has also at least one limitation that it precludes the assessment of the capability of the defendant to write texts. Both manuscripts have four authors and I assume that all or most of them contributed to writing. Unlike classical master thesis, here the personal writing style is probably lost, because other co-authors, with their names attached to the manuscript have definitely had more significant influence on the shape of the text unlike classical thesis written by the defendant under supervisor's surveillance. So, yes the English language and style is pretty fine, but I do not know how much if it is Eva Šochová style.

Regarding the quality of the manuscript drafts, I would like to state that I find them significant pieces of work that are worth to publish and I would like to congratulate Eva Šochová that she has gathered such large amount of the data during her master studies, I would like to stress that I have never seen a thesis that would be directly publishable in two regular papers, so I am amazed. The task of the reviewer is also to criticise, and so I will. My general criticism goes to the quality of the figures. They are generally too small and gave too small resolution that precludes reading branches support (this problem is most extreme in figure 2 of the second manuscript). I also do not understand, why the branch supports are given only in Bayesian posterior probabilities, if the ML analysis have been also performed. I would like to point also to the fact that although both papers should have extensive supplementary material, it is not attached to the thesis, although it is referred in the text. In the second paper, there as a link to the webpage of the journal, however, I was not able to find the material there.

Specific comments and questions to the thesis:

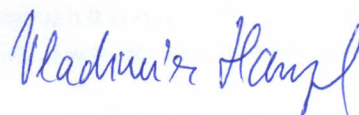
Manuscript 1

1. In figure 2 and related text you divide the *Arsenophonus* symbionts to obligate and facultative. How you know, which category the organisms belong to, if you have only 16S rRNA sequences?
2. Similarly in Figure 3 you divide *Sodalis* symbionts. As the criterion for the division you mention their host specificity, however, the number of sequences in the tree is rather low to validate the significance of the coevolution with the host. The obligate symbiont clade include sequences from three host species (two genera), the facultative branch of 5 species. The congruence with the host phylogeny in obligatory clade can be caused by chance. The incongruence with host phylogeny in the facultative clade is not obvious. The branching of *Crataerina* symbiont within *Ornithomya* is supported by invisible posterior probability.
3. Important finding of the manuscript is that the endosymbiosis in Hippoboscidae is very dynamic and symbionts are repeatedly exchanged. I wonder, what it the source of endosymbionts? Are they recruited from the intestinal bacteria? What is known about intestinal bacteria in Hippoboscidae?

Manucrypt 2

1. One of the main results of this manuscript are genomic sequences of two insect endosymbionts. The authors report that the genome of *A. ornithomyarum* contains also a plasmid that has been acquired by HGT from *Sodalis* bacteria present in the same environment. I would appreciate if these statements were better documented. Firstly, I do not understand, how authors know that the plasmid belong to *A. ornithomyarum*? The DNA sample from *Ornithomya* intestines contained at least three bacteria including one *Sodalis*-like, so the plasmid could belong to any of these. Secondly, I have not seen any evidence that the plasmid is related to *Sodalis* as they claim. I suppose they have performed phylogenetic analysis, but this is not included in the thesis.
2. The author mentions striking correlation between the level of endosymbiont genome reduction and the localisation of endosymbionts in a vacuole or freely in the cytoplasm? Vacuole localised symbionts in sap-feeding insects tend to be more reduced? She speculates that the reason is that they are more dependent on the host that is controlling the flow of nutrients through the vacuolar membrane. This is not for me a convincing hypothesis. I imagine that any restriction of the metabolite flow should force the endosymbiont to keep its own biochemical capacities and not to lose them and reduce the genome. Could you explain your reasoning a little bit more?
3. My final question is rather general. I wonder, if it is possible to recognise just from the genome sequence (without any other information), the lifestyle of the bacterium. Whether it is free-living, parasitic, endosymbiotic, insect endosymbiont (facultative vs. obligate). If any of this category can be determined from the genome itself, please explain briefly how.

Finally I would like to state that both the amount of delivered results and the formal quality clearly qualify the work for a very solid master thesis.



Vladimir Hampl

List of typos and small mistakes:

Page 16 line 32: typo *Glossina morsitans*

Page 32 line 6: the word "coverage" appears twice.

Review of Master thesis of Eva Šochová 'Evolution and genomics of symbionts in Hippoboscidae'

Even though it was the third thesis to review for me this semester, I was looking forward to read it and accepted to review it without a moment of hesitation. Personally, I find the topic of insect symbionts really interesting. I also like the way it's research is handled here and I already reviewed Bachelor thesis of Eva, so naturally I was curious about the progress during the past two years. And the progress it (at least to some extent) is! The presented thesis is in English and it contains two manuscripts in an advanced stage of completion. I'd call it a rather generous concept, which could still pass as a PhD thesis at some places quite easily. Still very uncommon around here, kudos to author and supervisor for such a boldness. While it follows the methodology of it's predecessor, it expands it in many aspects and shows a necessary progress.

On the other hand, this concept somehow blurs the contribution author of thesis. So, by giving compliments on the level of English the thesis is written with, I am actually not sure, whether I compliment the English of author, or her supervisor Filip Husník or other co-authors. The compliment is, by the way, deserved. Given the stage scientific career of author, the English is really good, although I recommend correction from native speaker or more seasoned author before submission. Some sentences still give away it's Czech structure and word order (I'll gladly give more details to authors if necessary) and there are several typos present (as in any thesis).

Both manuscripts are of comparable quality, I personally like better the second, but each could pass as a Master thesis on its own without substantial problems. The first part/draft compares the phylogenetic structure of hippoboscid flies to the diversity and phylogeny of their bacterial symbionts and aims to infer patterns of evolution of symbiosis in this insect group. It is apparent author performed lot of sequencing and lot of phylogenetic analyses.

However, their presentation is the main downside of the thesis in my view. The text contains four figures, the rest of results is deposited in Supplemental data. However, I have not obtained these in my copy of thesis, I had to ask for electronic version which was not accompanied with any details about what is shown. Also, I found the phylogenetic trees of really poor graphic quality. Supporting values are unreadable because of a small font used, color labeling of clades is (due to its transparency and bad choice of colors) ambiguous and requires lot of effort from the reader to understand and to follow the conclusions. Ironically, the exactly same point was my main critique of Eva's Bachelor thesis. Short Results are followed by a lengthy Discussion, which is detailed. Author shows deep knowledge of topic and works well with references to strengthen her own conclusions.

Other than that, I have just few minor points

page 14. - author mentions discrepancies between observed topology within Hippoboscoidea and already published results. She explains it by a weak phylogenetic signal. Is there any other possible explanation? She also mentions no EF for Hippoboscoidea in GenBank. Does that mean all the EF sequences originate from author?

page 15. author suggest LBA affecting observed topology. Apart from removal of problematic sequences, did she try to apply suitable phylogenetic methods to minimize the possibility? If so, which; if not, why?

page 19. phylogenetic methodology - host phylogeny: were concatenated dataset partitioned before analysis and coding sequences analyzed in nt or aa mode? What were the conditions of analysis for this dataset? Why author used PhyML instead of RAxML which is usually significantly faster, especially for nucleotides and performs better (just curious)?

The second manuscript describes genomes of two symbionts of avian louse flies and tries to reconstruct their metabolic potential from the perspective of symbiosis. This part of thesis seems to be more challenging, both on methodology and interpretation of results. Author apparently managed to deal with both problems. Introduction is nicely written and informative, but still maybe a bit long for a manuscript. I recommend to streamline it a bit. Results are concise, but the quality of Figure 2. and 3. again a bit disappointing and lacking. Discussion has similar quality as in manuscript #1, but I suggest structure it a bit. As I stated above, I like this manuscript better, however still have several comments/questions:

page 28 Introduction - author speculates on the possible differences between the blood of mammals and other vertebrates. Are there any data on blood composition available to support it?

Page 30 - maybe 23 genes should not justify the analysis to be called phylogenomic? :-). BTW, do the genomes of taxa included in analysis share only 23 single-copy genes or the number was reduced due for the purpose of feasibility of analysis? Why author used only the Gblocks, or why use it at all, did she try to manually edit the alignments? Also, the parameters of Gblocks analysis are missing.

Page 32 - author states *Arsenophonus* bacteria form a monophyletic clade. Short look at Fig.2 reveals it is not true, more over, from the text it is not clear, what method/software was used for a tree construction. Were there any differences between the topologies yielded by different methods, if so, did it affect the interpretation of results?

All in all, I can tell author did a tremendous job and made a big progress compared to her Bachelor thesis. However, there are several issues, mainly with the way how she present her results that are lacking and could still be improved. On the other hand, it would be ridiculous to punish her for that by lowering the marking, because her thesis is the best I reviewed so far. In spite of my objections to some minor issues I have no doubts it meets all the requirements for a successful defense, I suggest to award it with the grade **1** and I wish the author a good luck in her future scientific career.

České Budějovice, May 21st

Aleš Horák
BC CAS, Institute of Parasitology

