

Review on the Diploma thesis of Bc. Katarina Chalupecka

I read with great interest the Diploma thesis of Ms. Chalupecka, entitled "Comparison of the genetic variability of geographically distant populations of selected species of European marine gobies".

This study is examining the genetic variability of the populations of eight species of gobies, distributed mainly in Mediterranean Sea. The author has an extensive collection of samples (350 individuals) from eight countries, covering mainly the shores of Europe. The genetic variability is explored by analyzing one mitochondrial (cyt b) and one nuclear gene (S7). During her diploma work, Ms Chalupecka was trained on a variety of molecular methods (eg. DNA isolation, PCR, nucleic acid purification) and bioinformatic tools (eg. BEAST, DnaSP, Arlequin).

The thesis presents two major findings:

- a. surprisingly, seven out of eight species does not show any clear geographical structure of the populations, despite their extensive distribution (i see as single exception only *G. cruentatus*).
- b. cryptic diversity was identified in *Chromogobius zebratus*, which might consists of three distinct species.

Overall, despite some minor typos, the thesis is well written. The results seem solid and convincing, the discussion is straightforward, the figures and tables are well prepared and informative. Therefore, I recommend the thesis to be accepted.

I have only two comments, which need the author's attention:

**Comment 1:** In pages 34-35 and 80-81, the author show more haplotypes and genetic diversity in the nuclear S7 gene than in the mtDNA gene cytb, in *C. liechtensteini*. Such finding is quite surprisingly, as it is usually expected that nuclear genes evolve much slower than mtDNA and it is contradictory to the findings in other species of gobies. Assuming that the mutation rates are probably similar in all gobies for both genes, the author speculates that this pattern could be explained by a bottleneck event and subsequent expansion of the population(s), from a limited number of female individuals. I think that this might be the most interesting finding of the study and deserves further investigation. This scenario could be easily tested by sequencing one or two additional nuclear markers (eg. RAG1 or rhodopsin) and check if more nuclear genes show the same pattern. However, there are more plausible scenarios than

bottleneck event and subsequent expansion. The authors identified the alleles of the S7 per individual by bioinformatic tools (Phase), therefore, it is possible that some of the haplotypes might be artifacts. In fact, the author should take into consideration that a. it is possible that individuals of this species might have duplications of S7 gene, which Phase does not sort, b. the primers might amplify a pseudogene in some cases, c. the individuals are not diploid but polyploid (such cases are often reported in fish species). In order to solve this problem, it would be preferable "cloning" the PCR products of S7 and sequence multiple clones, than using Phase software.

**Comment 2:** I am not convinced on the proper use of GENELAND software for this study. This software estimates "*population structure in form of systematic variation of allele frequency that can be detected from departure from Hardy-Weinberg and linkage equilibrium...* It implements several models that can make use of both geographic and genetic informations to estimate the number of populations in a dataset and delineate their spatial organisation" (see <http://www2.imm.dtu.dk/~gigu/Geneland>). The disadvantages in the system of interest (gobies) are plenty, including: a. relative few localities for this software, b. the localities have huge distances between them and are not randomly selected across the species geographical distribution, c. localities are missing for the South Mediterranean region and d. the studied gobies are mainly distributed in relative shallow waters around shores and not across large areas of deep waters, therefore, the landscape is not uniformly colonized. Furthermore, I assume that a requirement to use this software, where the population borders are defined from "*allele frequency that can be detected from departure from Hardy-Weinberg and linkage equilibrium*" is that your populations should have already some structure and not to be panmictic, which it is not fulfilled in six out of eight studied species according to AMOVA and TCS networks.

Sincerely yours,

Michail Rovatsos