

## lab8\_questions\_and\_maj\_answers

**Author: Mohammed Al-Jaff**

---

### Recap

**"We will also expect you to remind us briefly of the question you worked with, and list which species you chose to include and why, so take a moment to prepare that too."**

Chosen question:

**4) Are salamanders more closely related to frogs than to lizards?**

Chosen species for each of the three groups:

Frogs	Salamanders	Lizards
<i>Pelophylax cretensis</i> (Greek marsh frog)	<i>Necturus beyeri</i> (Gulf Coast waterdog)	<i>Lacerta agilis</i> (Sand lizard)
<i>Rana amurensis</i>	<i>Ambystoma texanum</i>	<i>Scincella modesta</i>
<i>Xenopus borealis</i>	<i>Andrias japonicus</i>	<i>Plestiodon elegans</i>
<i>Rana temporaria</i> (common frog)	<i>Batrachoseps nigriventris</i> (black-bellied slender salamander)	<i>Iguana iguana</i> (Common green iguana)
<i>Amolops ricketti</i>	<i>Hynobius leechii</i> (Gensan salamander)	<i>Leiocephalus personatus</i> (Haitian curlytail lizard)
	<i>Triturus carnifex</i>	<i>Podarcis muralis</i> (Common wall lizard)
		<i>Gekko japonicus</i>

Outgroup Species, both original faulty outgroup choice and subsequent proper outgroup choice:

<b>Sharks (Proper Outgroup)</b>	<b>Primates (initial faulty)</b>
<i>Carcharhinus obscurus</i> (dusky shark)	<i>Pan troglodytes</i> (chimpanzee)
<i>Carcharias taurus</i> (sand tiger shark)	<i>Gorilla gorilla gorilla</i> (western lowland gorilla)
<i>Cetorhinus maximus</i> (Basking shark)	<i>Pongo pygmaeus</i> (Bornean orangutan)
<i>Triacodon obesus</i>	
<i>Orectolobus japonicus</i> (Japanese wobbegong)	

For NCBI links to each organism mitochondria genome see my answers/report for lab6. Please also refer to the same report for a more detailed explanation of our species/group/out-group strategy choice. In short, our choice basically was based on availability of genomes and having enough diversity/representation within each of the above groups.

In the tree figures mentioned below and attached in the submission, I have chosen to label the sequence of each species such that the first letter of each header for each group begins with an uppercase letter of the group name, ie All frog sequences begin with 'F', all salamander species begin with 'S', all "lizard" species begin with 'L' and all the species in the outgroup shark begin with 'O1'. You can see in the 3 accompanying tree pdfs that in both the cytB runs and in the whole-mitochondrial genome runs that all species within each group ended up in their correct respective clades.

---

**Q: What is the answer to the question given these trees? & Q: Do the different trees give the same answer? If not, what could be the explanation?**

**Ans:** The two trees I ended up with gave conflicting answers to the project question. The phylog tree obtained using only the *CytB* gene sequences of all the species points at salamanders being closer related to frog than lizards. The opposite conclusion is supported by the tree obtained using whole mitochondrial genomes of all the species. Here it turned out that the maximum likelihood tree was one where lizards and salamanders were more closely related pair. [all all cases the consensus tree agreed with the ml tree in each respective conclusion]

--

**Q: Which tree(s) do you think are most supported? What does the bootstrap analysis tell?**

**Ans:** I can make arguments for both positions. For the *cytB*-based phylogenies, we can argue that we are dealing with a single highly conserved gene that is relatively easy to align 'correctly' in a multi-sequence alignment which would result in an evolutionary distance matrix with higher confidence which in turn would lead helping at getting a more "unlikely" tree reconstruction. Not only that, but this phylogeny is in agreement with the current consensus of the 'true' phylogeny which states that salamander and frogs are the closer related pair [source: **Meyer, Axel, and Rafael Zardoya. "Recent Advances in the (Molecular) Phylogeny of Vertebrates." *Annual Review of Ecology, Evolution, and Systematics* 34, no. 1 (November 2003): 311-38.**]

On the other hand, one could argue that the mitochondrial genome is packed with more or less moderate to highly conserved genes and the collection of all of these together would constitute a more robust 'instrument' and measure of evolutionary divergence. And as such whole mitochondrial genome alignment would lead to a more 'accurate' evolutionary distance matrix which would provide the foundation of any tree reconstruction analysis. But then again, in our particular case, we don't seem to have ended up at the 'right' answer/conclusion.

**Q: Are the substitution models different for cytB and the entire mitochondria? If yes, what could be an explanation?**

**Ans:**s Yes, we obtained different 'best-fit' substitution models for cytB and mitochondrial genome runs. This makes sense given that for the cytB case, we are dealing with a single core/vital gene with its own tolerance to various mutation types across its sequences while in the mitochondrial case, we are dealing with a whole bunch of mixed 'conserved' genes in addition to inter-genic regions with neutral selection to mutations in many parts.

---

**Q: Do you think your choice of species (including the outgroup) was appropriate to answer the question? If you were to redo the analysis, would you change something?**

I don't think my 'Shark' outgroup choice was problematic because it fulfilled the criteria of being a good outgroup (being a proper outgroup while being "close" enough so that a good alignment to the rest of the sequences would be achievable). Nor do I think I messed up by including my 'primate' group in the dataset because i) you could simply view this as a group of sequences that would have been involved in another hypothetical project question. ii) I redid the whole multiple sequence alignment + phylogenetic reconstruction steps with the same dataset excluding the primate groups and ended up with the same results that the whole-mitochondrial-genome based phylogeny still indicated that lizards are closer to salamanders. So the problem must have been in either not enough sequences for each group, OR using faster but less precise algorithms for the alignment and phylogeny steps OR using whole mitochondrial genomes instead of concatenated sequences of core conserved genes/regions. If I were to redo the whole project I might address one or all of these and see if that would result in a 'better' result that agreed with the consensus phylogeny.