THE RED PANDA AND CSERHATI (13): THE BMC GENOMICS ARTICLE

In the BMC Genomics article, Cserhati uses two datasets and two types of methods to find the best classification, the phylogenetic placement, of the red panda. The natural classification of the species represents their relationship, and relationships between species and families are always interesting.

The two types of methods used by Cserhati are clustering and phylogenetic methods. The two data sets are Whole Genome K-mer Signatures of 28 species, and mtDNA sequences of 52 species.

1   Clustering

Cserhati is looking for the "phylogenetic placement" of the red panda (as he says in the second sentence of his abstract) and then clustering is not a useful method. Clustering is not a phylogenetic method and does not provide a phylogeny, and therefore no "phylogenetic placement". Clustering cannot be used to conclude monophyly of a group or groups. Cserhati says in Methods: "*...using clustering algorithms to detect monophyletic groups"*, which is wrong, as clustering is not able to detect monophyly.

The discussion of the results of the WGKS data states:

> *Ailurus fulgens* clearly clusters together with the mustelids,

> Based on this evidence ["*Ailurus fulgens* clearly clusters together with the mustelid"], *Ailurus fulgens* would belong to mustelids as a monophyletic group.

The first sentence is not controversial but not interesting for the classification of the red panda *Ailurus fulgens.* The second sentence is wrong in two ways. First, monophyly cannot be derived from clustering. Second, "monophyly" and "belonging to the same family" are not the same thing.

2   Phylogenetic methods on the two datasets.

2.1   WGKS

One phylogenetic analysis method, UPGMA, was applied to the WGKS data set with 28 species.

The UPGMA phylogenetic tree mainly shows that WGKS is not very useful for phylogeny (BMC Genomics Figure 2). The very large differences, those between the cat family, bear family and the superfamily Musteloidea, are found correctly. The split found within the Musteloidea is among the possibilities, given all other studies on the red panda. Within the marten family, the weasels and the otters are intermingled. Within the cat family the 'big cats' and the 'little cats' are intermingled. For the more related species, the UPGMA analyses on WGKS gives garbage. In the Abstract, Cserhati writes:

> (*WGKS) Being a genomics-based algorithm, it also reduces stochastic error to a minimum*.

Judging by his results in figure 2, this is not correct. Stochastic error is high.

2.1 mtDNA

On the 52-species mtDNA dataset, Cserhati applied three phylogenetic methods.

The UPGMA method gives a placement of marten family and red panda as sister group; marten family Mustelidae and Ailuridae as sistergroups are not found anywhere in the literature, in the presence of species of the raccoon family.

Cserhati interprets the phylogenetic trees from the NJ analysis and the ML analysis according to the layout, showing a total lack of understanding of phylogenetic trees. Surely it should have alarmed Cserhati that the superfamily Musteloidea did not appear in his interpretation.

With the bear family as outgroup both NJ and ML analyses yield the same pattern within the superfamily Musteloidea: the first split is that of red panda Ailuridae against skunks + raccoons + marten family, the second split is that of skunks against raccoons + marten family. The raccoons and the marten family appear as sister groups. This is a well-known relationship pattern within the superfamily Musteloidea, found in a number of scientific papers.

Cserhati appears sloppy in his presentation of the results for the mtDNA dataset. The Abstract states:

> *A Maximum Likelihood tree suggests that* A. fulgens *and Ursidae form a monophyletic group, although the bootstrap value is weak.*

Ailurus fulgens and Ursidae a monophyletic group is Cserhati's interpretation of the NJ phylogenetic tree, not the ML tree. He mixes up his own results.

3   It is striking that in Cserhati's analysis of mtDNA the word 'outgroup' never appears.

Even more striking is that the idea 'sistergroup' does not occur at all. Two sistergroups are monophyletic. Where Cserhati wants to use 'monophyly' to argue that the red panda belongs to the marten family Mustelidae, that monophyly shows a sistergroup relationship between the families Ailuridae and Mustelidae. Monophyly does not mean belonging to the same family.

4   In the Abstract Cserhati says:

> *Conclusions: The main conclusion that we can draw from this study is that on a whole genome level <u>Ailurus fulgens</u> possibly belongs to the mustelid clade,*

This has no basis. It derives from the WGKS clustering, but 'cluster' and 'clade' are confused.

5   The introduction is teeming with errors. Almost all articles on phylogeny are cited incorrectly.

6   As I wrote on Panda's Thumb :

"One problem: how did a paper as bad as this ever get through review and published?"

 ***

 Cserhati, M., 2021, A tail of two pandas – whole genome k-mer signature analysis of the red panda (Ailurus fulgens) and the Giant panda (Ailuropoda melanoleuca), BMC Genomics 22: 228

http://pandasthumb.org/archives/2022/12/a-tale-of-two-papers.html