THE RED PANDA AND CSERHATI (7): WGKS IS NOT SUITABLE FOR PHYLOGENY

Cserhati prefers use of the entire genome for species classification, rather than a number of well-characterized genes. The method used by Cserhati to characterize the whole genome is Whole Genome K-mer Signature, abbreviated as WGKS. Cserhati clearly thinks that this WGKS method can be used to arrive at the proper classification of the red panda.

> *While the WGKS algorithm may not be a sensu stricto phylogenetic algorithm, it can still be used to classify species, based on their WGS into different groups. (BMC Genomics)*

Cserhati uses two techniques for red panda and giant panda classification, a phylogenetic tree and clustering.

This post is about the phylogenetic tree; clustering will be discussed later.

Cserhati uses a WGKS data set of 28 species

> *To this end, the Whole Genome Kmer Signature (WGKS) algorithm [15] is used to analyze the genomes of five bear species, eleven cat species and ten species from the family Mustelidae (weasels, otters, martens, and badgers), Spilogala gracilis, a mephitid species, as well as the red panda Ailurus fulgens, making 28 species in total.*

The five species of the bear family are the giant panda and four species of the genus *Ursus*. It is clear here that Cserhati takes the classification of the giant panda with the bears for granted.

The 28 species are listed with their scientific name and their English name in the following table:

familie	subfamilie	soort	nederlandse naam
Mustelidae		*Mustela erminea*	Ermine
Mustelidae		*Mellivora capensis*	Honey badger
Mustelidae	otter	*Pteronura brasiliensis*	Giant otter***
Mustelidae	otter	*Enhydra lutris*	Sea otter

Mustelidae		*Taxidea taxus*	American badger
Mustelidae		*Neovison vison*	American mink
Mustelidae	otter	*Lontra canadensis*	Northern river otter
Mustelidae		*Mustela putorius furo*	Ferret
Mustelidae		*Gulo gulo*	Wolverine
Mustelidae	otter	*Lutra lutra*	European otter
Ailuridae		*Ailurus fulgens*	Red panda
Mephitidae		*Spilogale gracilis*	Western spotted skunk
Ursidae		*Ursus thibetanus*	Asiatic black bear
Ursidae		*Ursus arctos*	Brown beer
Ursidae		*Ursus americanus*	Grizzly
Ursidae		*Ursus maritis*	Polar bear
Ursidae		*Ailuropoda melanoleuca*	Giant panda
Felidae		*Lynx canadensis*	Canadian lynx
Felidae		*Felis catus*	House cat
Felidae		*Puma concolor*	Puma
Felidae		*Lynx pardinus*	Iberian lynx
Felidae		*Prionailurus bengalensis*	Leopard cat
Felidae		*Panthera onca*	Jaguar
Felidae		*Panthera pardus*	Panther
Felidae		*Panthera leo*	Lion
Felidae		*Acinonyx jubatus*	Cheetah
Felidae		*Panthera tigris*	Tiger
Felidae		*Felis nigripes*	Black footed cat

Cserhati uses the WGKS data to create a phylogenetic tree with the UPGMA method. UPGMA is the simplest and oldest method to construct a phylogenetic tree from sequence data. The major drawback of UPGMA is that this method assumes that the rate of change in the sequence is the same over time and across all lines, ie over the entire phylogenetic tree. That is by no means always the case, and UPGMA is seldomly used anymore.

In his figure 2 (here figure 1) Cserhati gives the UPGMA phylogenetic tree for his 28 species. The layout in the figure has the same order of species from top to bottom as in the table above.
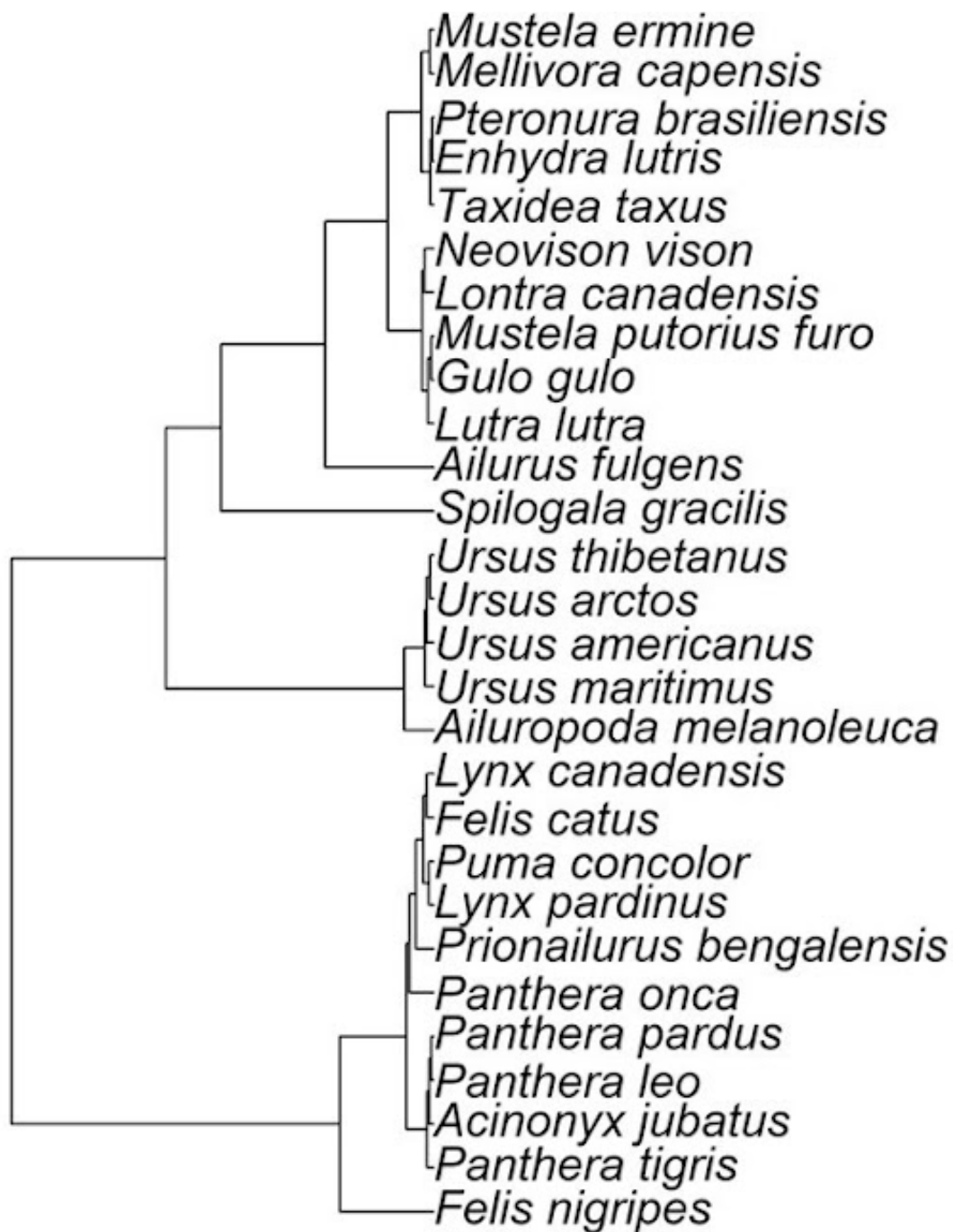
Figure 1. Phylogeny by UPGMA methode, WGKS data: Figure 2 Cserhati BMC Genomics.Horizontal line length corresponds to differnce found.

We clearly see three major groups in Figure 1. The first split is the cats Felidae against all other species, the second split is the bears Ursidae against the Musteloidea. Within the Musteloidea we have the Mephitidae splitting off first, and a sister group relationship of the red panda Ailuridae with the Mustelidae. This corresponds to the phylogenetic tree of Law et al (2018) (in blog post Cserhati 1), in the absence of the raccoon family Procyonidae.

No problems with the major groups, but problems surface in the phylogenetic trees for the families
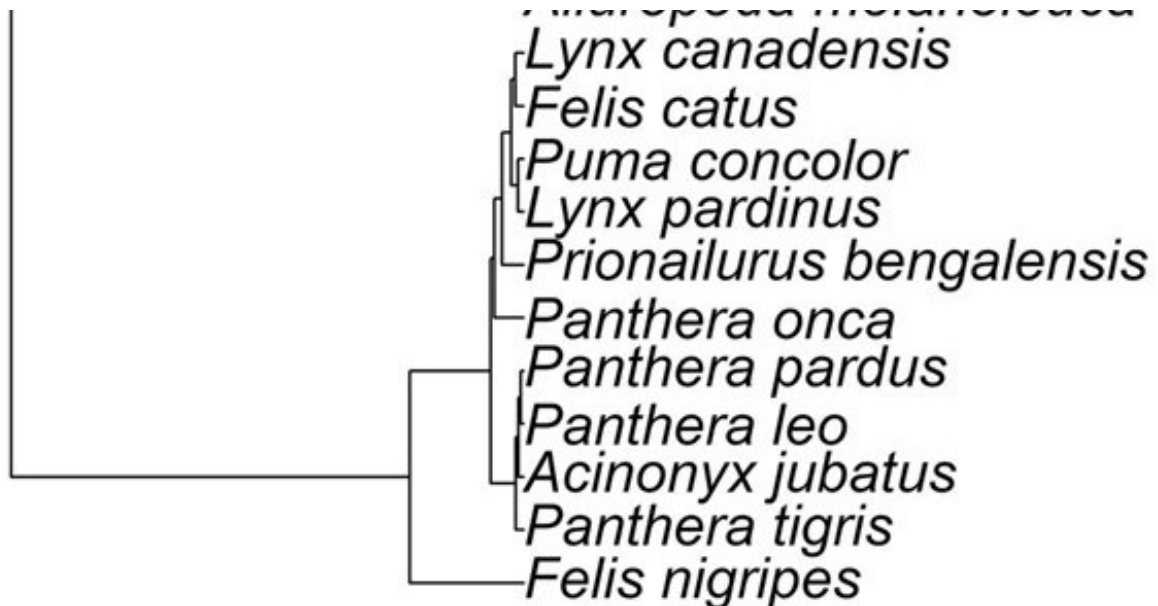
Let's look at the cats first:

Figure 2 detail of Figure 1, the cats

The phylogeny in Figures 1 and 2 indicates that the black-footed cat *Felis nigripes* is equally related to all other cat species: equally related to the lion, the lynx as to the domestic cat *Felis catus;* equally related to the big cats as to a cat species from the same genus *Felis*. That can't be right.

Moreover tiger, lion and panther are grouped together with the cheetah. That can't be right: the cheetah *Acinonyx* has never been counted among the 'big cats'. The jaguar *Panthera onca*, on the other hand, is here separated from the other 'big cats' of the genus *Panthera*. The Canadian lynx gets the domestic cat as a closest relative, but the Iberian lynx gets the puma as closest relative. Something is going very wrong here.

The species of the mustelid family Mustelidae are messed up too, as much as the cat species:
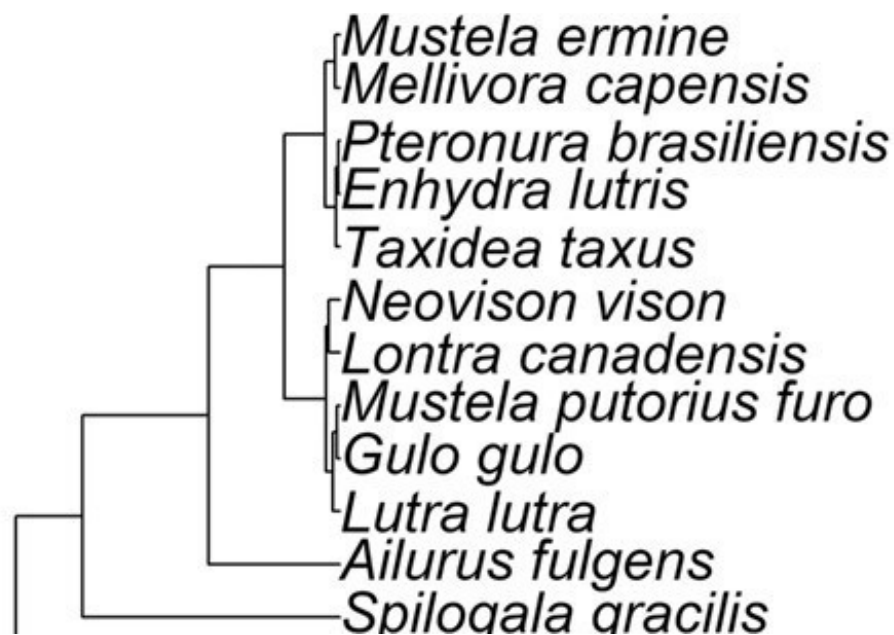
Figure 3. Detail Figure 1, superfamily Musteloidea

The phylogeny in Figures 1 and 3 indicates that the European otter *Lutra lutra* is more closely related to the wolverine and the ferret than to the other three otters. The ferret and ermine, both from the genus *Mustela*, are thrown apart in the first split within the family Mustelidae. One does not need to have a background in biology to see that the positions of the otters and weasels are messed up. The otters and *Mustela* clearly show that this phylogenetic tree cannot represent the relationships between the species in a correct way.

Altogether, the UPGMA phylogenetic tree on Whole-Genome K-mer Signatures shows well-known results among the major pattern of families and superfamilies. On a coarse level WGKS classifieds correctly. Within a family, WGKS cannot be used to assess relatedness. On a more detailed level, WGKS gives junk. It is impossible to say where the transition lies between 'coarse is correct' and 'fine is junk'.

What does Cserhati say about his Figure 2?

> *Based on this evidence, A. fulgens would belong to mustelids as a monophyletic group. ....This can also be seen well in Fig. 2, which shows the UPGMA-based phylogenetic tree for the 28 species in the whole genome analysis.*

Not so. Note the long horizontal line length between *Ailurus* and the species of the Mustelidae family In fact, Cserhati's Figure 2 shows the red panda as the sister group of the family Mustelidae; it does not show the red panda to belong to that family.

******

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

https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-07531-3

https://en.wikipedia.org/wiki/UPGMA

Law, C. J.; Slater, G. J. & Mehta, R. S. (2018). Lineage Diversity and Size Disparity in Musteloidea: Testing Patterns of Adaptive Radiation Using Molecular and Fossil-Based Methods. Systematic Biology. 67: 127–144.