THE RED PANDA AND CSERHATI (15): DATA AND RESULTS IN THE CRSQ ARTIKEL

In his Creation Research Society Quarterly article, Cserhati uses three types of data to find the red panda's place among other species: Whole Genome K-mer Signature, mitochondrial DNA sequence, and amino acid sequence in the protein cytochrome-b.

The first two sets of data, WGKS and mtDNA, are the same as in his BMC Genomics article. For the cytochrome-b amino acid sequence, Cserhati uses new data. Data on the following numbers of species/subspecies are used:

family	WGKS	mtDNA	aa cytochrome-b
cats Felidae	11	-	25
bears Ursidae	5	15	5
Martens Mustelidae	10	30	17
raccoons Procyonidae	0	2	1
skunks Mephitidae	1	3	1
red panda Ailuridae	1	2	2
	28	52	51

The raccoon family is absent from the WGKS data, although Cserhati could have collected data for two species. The paper by Tsuchiya et al in *Genome Biology and Evolution* detailing the kinkajou and raccoon genomes was published in January 2021. This would have been in time for use in the CRSQ article in Fall 2021.

WGKS

From the WGKS data, Cserhati calculates a correlation matrix between the species. This is the same correlation matrix as in the BMC Genomics article, as

evidenced by the two Additional Files. Cserhati provides heatmaps of the correlation matrix in both articles: the same figure, but colored differently.
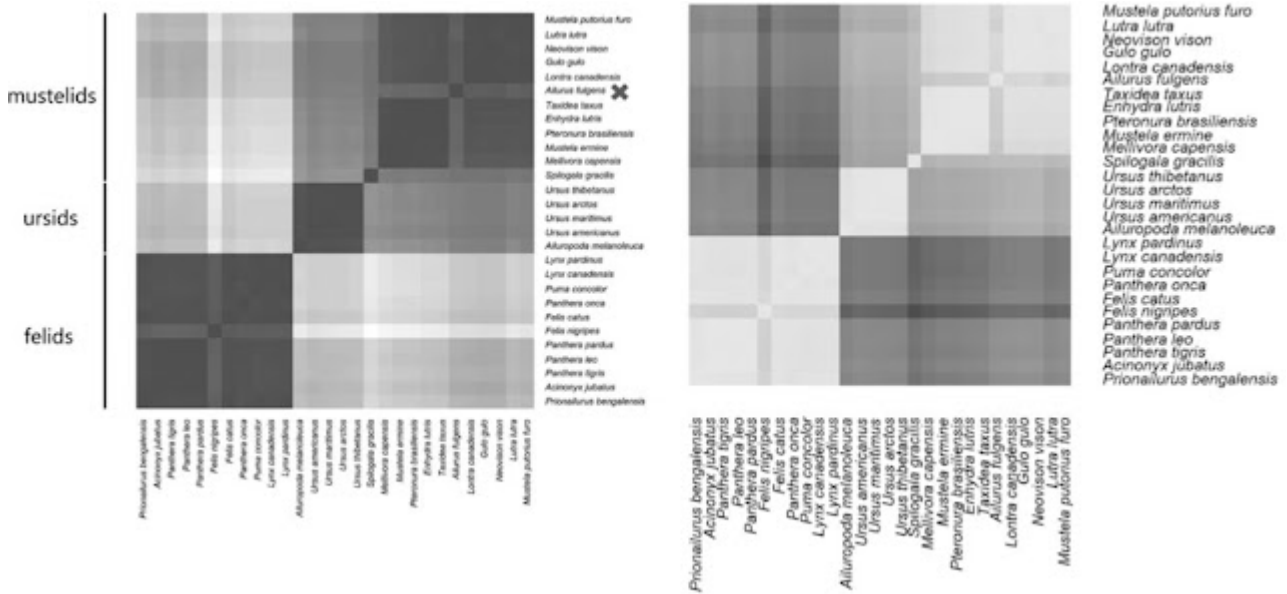

Figure 1. Heatmaps of the correlations on the WGKS data: left as in CRSQ article, right as in BMC Genomics article. The gray scale is reversed, otherwise it is the same figure.
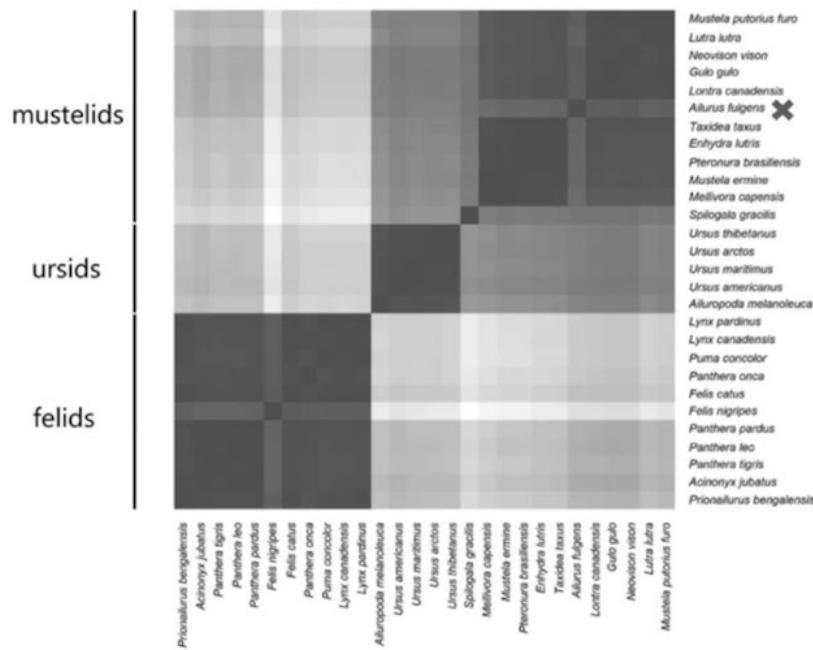

Figure 2: Heatmap indicating the size of the correlations in WGKS between the species. Darker is more similar. The order of species on the axes is the same from top to bottom as from right to left; the southwest-northeast diagonal gives correlation 1, the correlation of the species with itself. The families are indicated on the left: from top to bottom marten family, bear family, cat family. Note that here both the western spotted skunk *Spilogale gracilis* and the red panda *Ailurus fulgens* are reckoned to be mustelids, to belong to the marten family Mustelidae rather than to the superfamily Musteloidea. Figure 2 CRSQ 2021.

The location of the red panda *Ailurus fulgens* is indicated on the right, in the middle of the mustelids. This leads to a kind of 'Danish flag' pattern. The only species of the skunk family Mephitidae in these dates, *Spilogale gracilis*, can be found between the bears and the martens, the block next to the 'Danish flag'.

In this figure, Cserhati indicates families: felids, the cat family Felidae, ursids, the bear family Ursidae, and mustelids, the marten family Mustelidae. Apparently, the red panda and *Spilogale* are counted in the marten family in the figure.

In the text, Cserhati however says that the three clusters found in the WGKS data are the cats, the bears and the superfamily Musteloidea:

> *… felids covering the family Felidae (eleven species), ursids covering the family Ursidae (five species), and musteloids, a superfamily including Mephitidae and Mustelidae (twelve species).*

Two points about Cserhati's 'three clusters' result deserve notice.

The first point of interest is that Cserhati arrives at three clusters in this CRSQ article, but arrives at four clusters with exactly the same results in the BMC Genomics article . In the BMC Genomics article, the skunk *Spilogale gracilis* is set apart as the fourth cluster. Here in the CRSQ article, *Spilogale gracilis* is included in the large cluster Musteloidea. Cserhati gives silhouette plots for 3 clusters and for 4 clusters: four clusters if *Spilogale* is held to be a separate cluster. The zero silhouette width for *Spilogale* indicates that clustering in the adjacent cluster is better. Therefore, the silhouette plots show better clustering with three clusters.
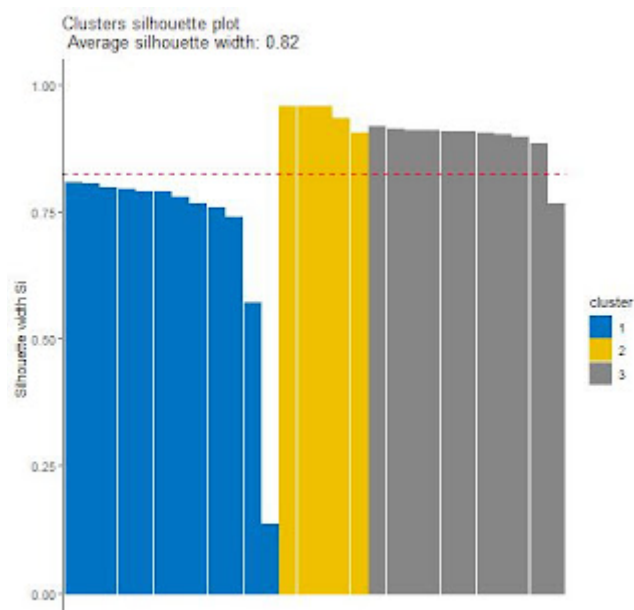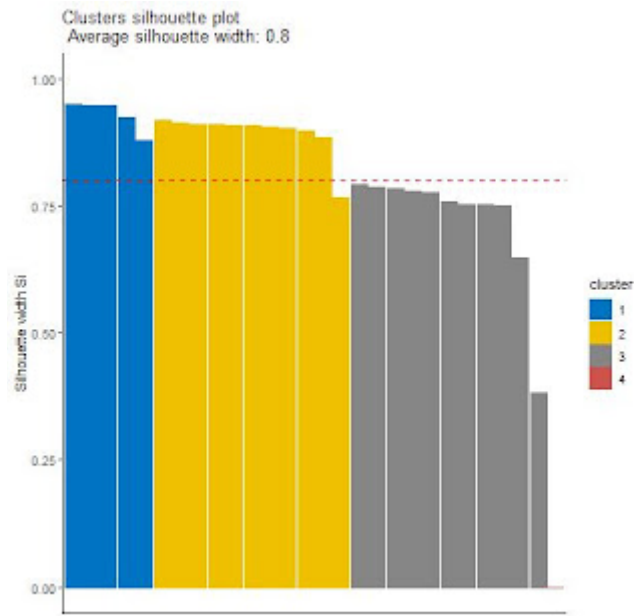
Figure 3. Silhouette plots for 3 clusters (above, CRSQ supplementary figure 1) and for 4 clusters (below CRSQ supplementary figure 2). Cserhati gives no description for the x-axis. The five species of the bear family are yellow above, blue below. The eleven species of the cat family above gray and below yellow. The twelve species of the superfamily Musteloidea are blue above and grey/red below.

The second point of interest is that Cserhati classifies the red panda in the marten family Mustelidae - something he should have shown, but here assumes without any justification. The red panda is different from all mustelid species: as can clearly be seen in the heatmap
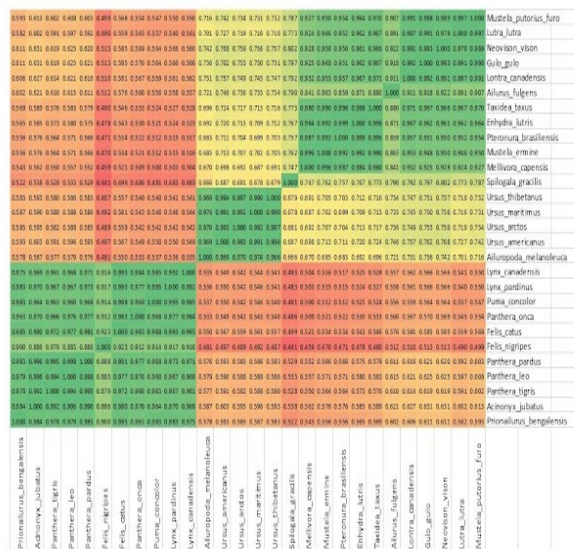

Figure 4. Colored WGKS heatmap from CRSQ paper: greener is higher correlation, redder is lower correlation. Minimum Correlation: .44. This heatmap has exactly the same correlations as black and white Figure 2 here.

Figures 2 and 4 clearly show the major difference in the correlations is between the cats on the bottom left and all other species. To clarify differences within the superfamily Musteloidea, a heatmap with just the correlations within the Musteloidea is called for.
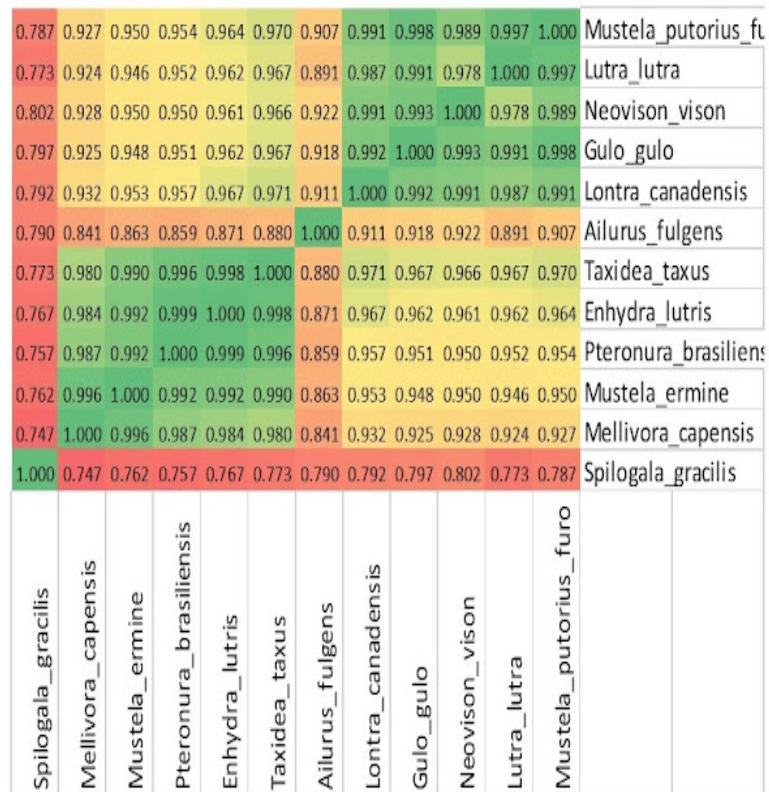

Figure 5. Heatmap of the Musteloidea, no cats and bears. The order of the species is the same as in Figure 2 and Figure 4. Now the lower correlation between the red panda *Ailurus fulgens* and the species of the marten family is clearly visible.

The difference is statistically significant (blog post 9). The red panda may cluster within the superfamily Musteloidea, but it is not belong to the family Mustelidae.

mtDNA sequences

Cserhati calculates a correlation matrix between the species from the mitochondrial DNA sequence data. This is the same correlation matrix as in the *BMC Genomics* article. Cserhati provides a heat map of the correlation matrix: the same heatmap as in *BMC Genomics,* but colored differently and with the species in different order.
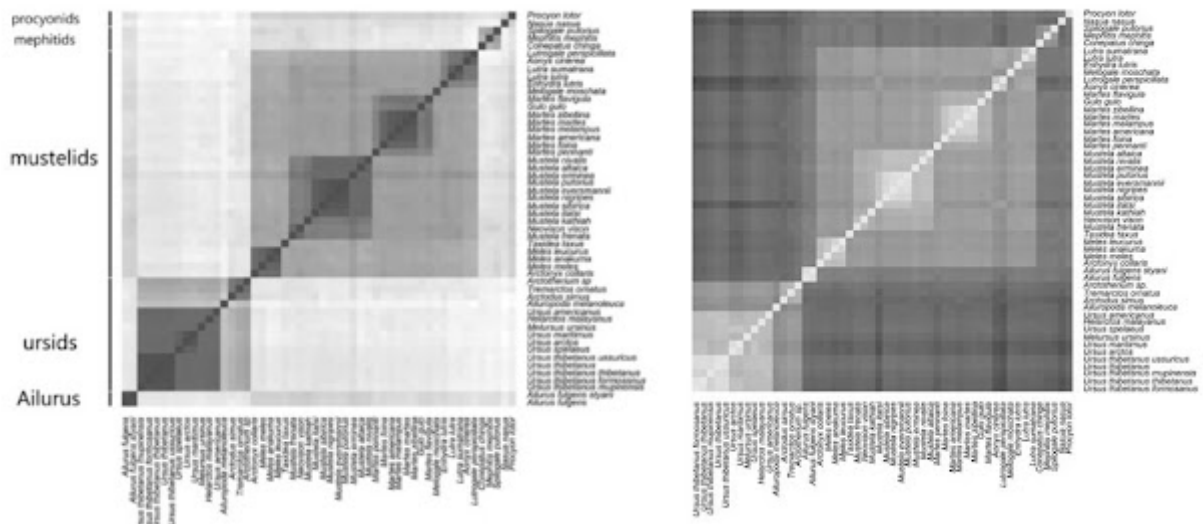
Figure 6. Heatmaps of the correlations on the mtDNA data: left as in CRSQ article, right as in BMC Genomics article. The gray scale is reversed. The two subspecies of the red panda are placed at the very bottom left in the CRSQ article, and can be found in the BMC Genomics article as the small light block between the two larger light blocks.
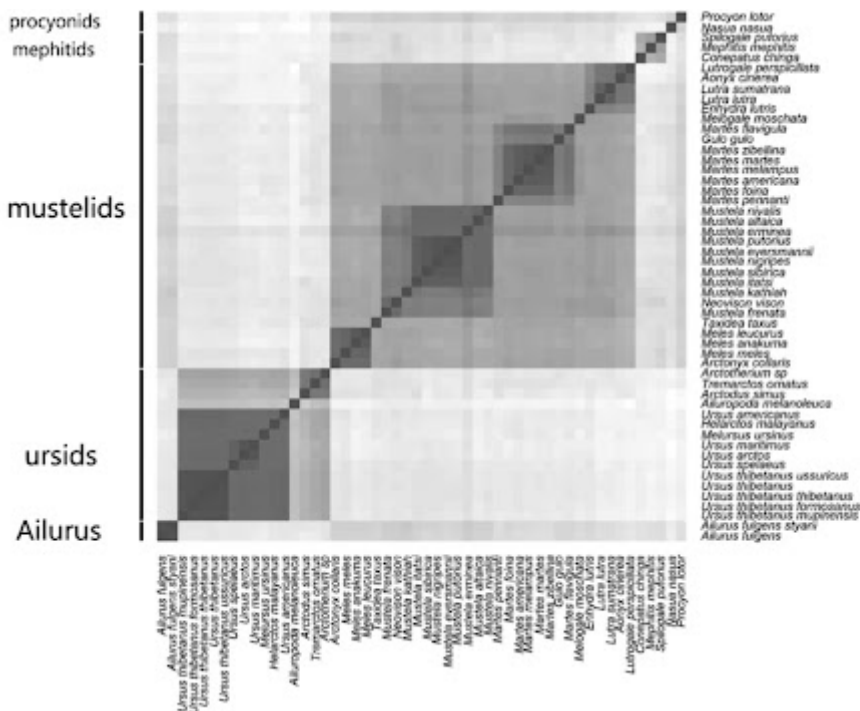

Figure 7:
Heatmap of the correlations in mtDNA sequence between species. Darker is more similar. The order of species on the axes is the same from top to bottom as from right to left; on the diagonal southwest -northeast is the correlation with the species itself. The two red panda subspecies are at the bottom left. Figure 3 CRSQ 2021

Cserhati comments:

In this heatmap we can see two main, large clusters as well as three smaller ones.

The bears, ursids, and the marten family, mustelids, will presumably be meant by 'the two large clusters'. The interpretation with two large clusters and three small clusters is made more intuitive by the placement of the two red panda subspecies in the lower left (*Ailurus*) and the placement of the two raccoons (procyonids) and the three skunks (mephitids) in the upper right. As a consequence, it is difficult to see the correlations of the red panda with the mustelids, mephitids and procyonids: light gray along the lower edge, indicating the two red panda subspecies are not well placed in this figure.

The correlations in the heatmap can be found in Additonal File 2 of the article. I've changed the order of the families along the axes, and brightened up the correlation matrix for a clearer red-green heatmap.
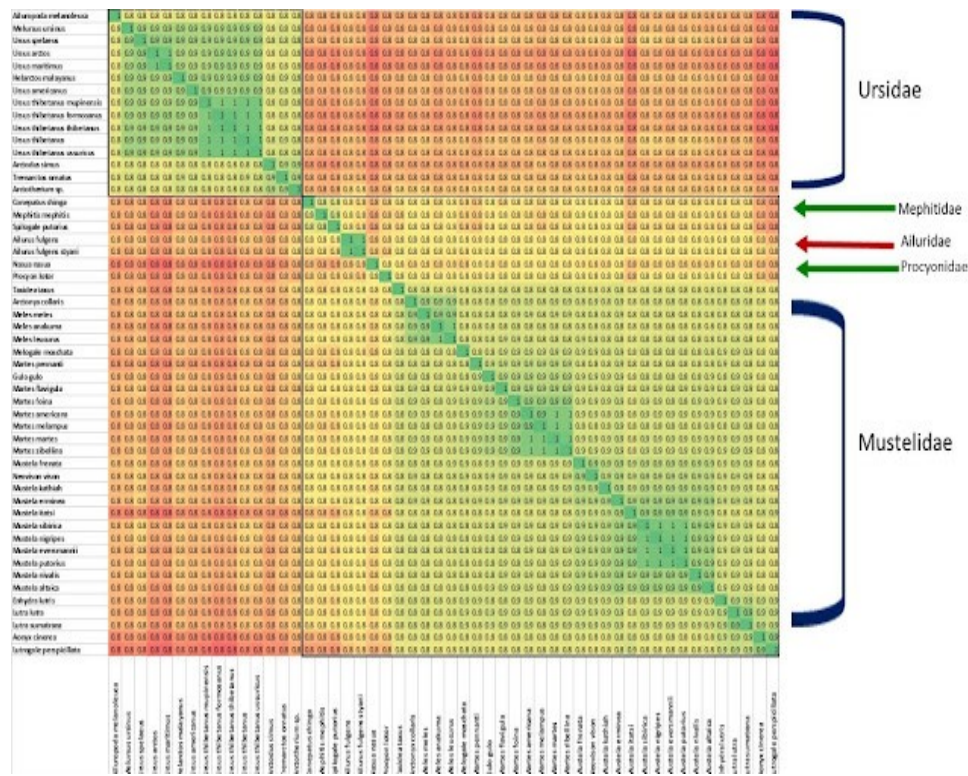

Figure 8. Heatmap indicating magnitude of correlations in mtDNA sequence between species. Greener is more similarity, redder is more difference. Minimum Correlation: .75. The order of species on the axes is the same from top to bottom as from right to left; on the diagonal northwest-southeast is the correlation with the species itself. Figure 3 CRSQ 2021 with reordering of the families and in different colors.

The resulting heatmap provides a clearer view. There are two main groups, one with the bear family Ursidae on the top left and one with the superfamily Musteloidea on the bottom right (including the families Mephitidae skunks,

Ailuridae red panda, and Procyonidae raccoons). Within the Musteloidea more variation is found than within the bears. The Musteloidea show families and subfamilies.
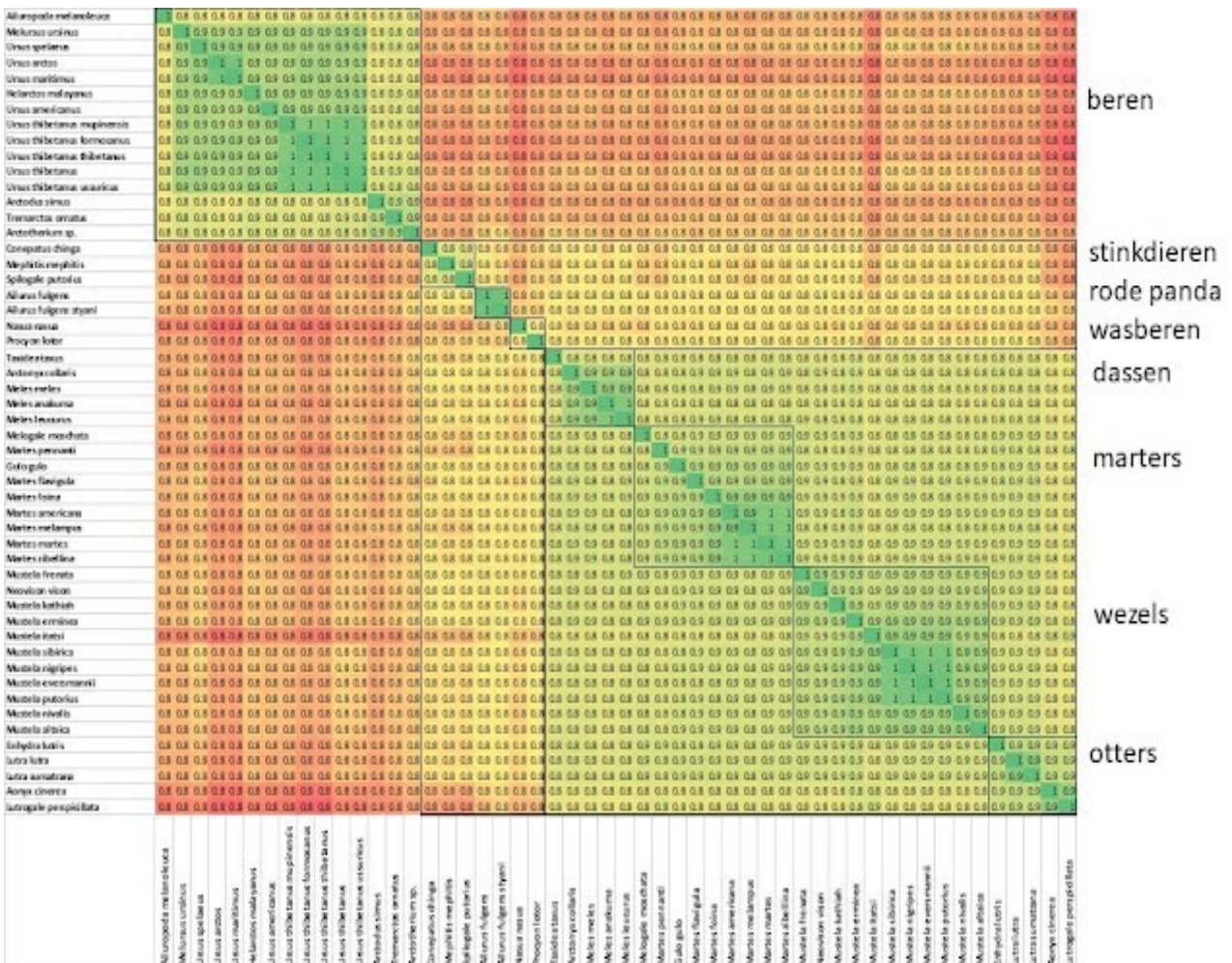

Figure 9. Heatmap from figure 8 with families and subfamilies within the Musteloidea indicated. Beren = bears; stinkdieren -= skunks; rode panda = red panda; wasberen = raccoons; marters = martens; wezels= weasels; otters =otters.

Some differences show up between the mustelid family Mustelidae and the other three families in the superfamily Musteloidea. It's a little more orange in those lanes than in the square. The differences can be put into focus.

A heat map of the superfamily Musteloidea on its own emphasizes the differences. In such a heatmap, the major differences between bears and Musteloidea have disappeared. The differences within the Musteloidea become visible with continued heatmap magnification and focussing.
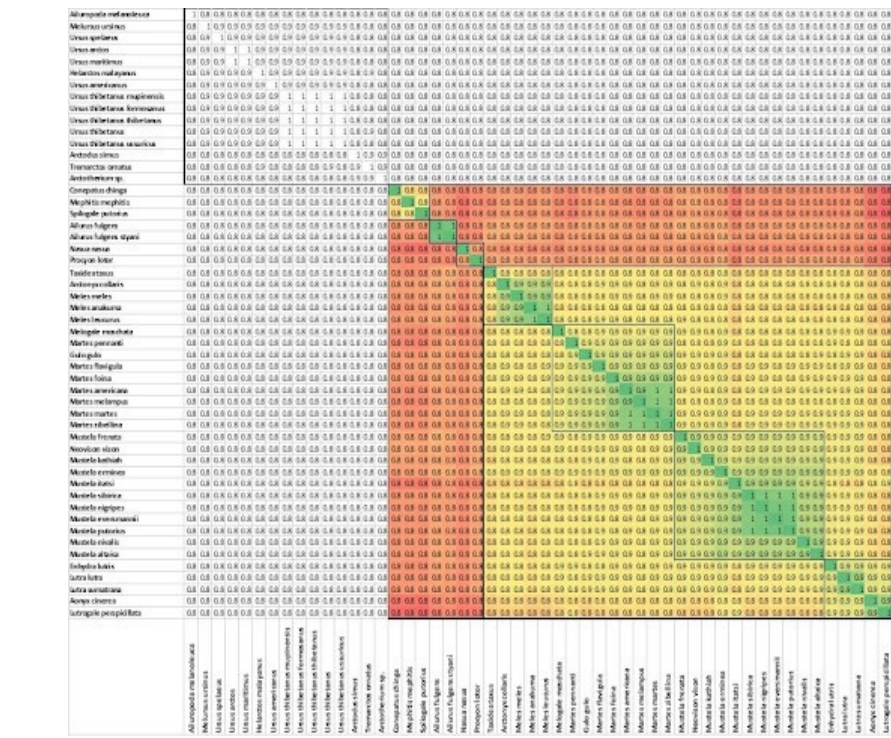
Figure 10 Identical correlation matrix as in the figures 7, 8 en 9 emphasizing the differencesin the correlations within the Musteloidea. Beren = bears; stinkdieren -= skunks; rode panda = red panda; wasberen = raccoons; marters = martens; wezels= weasels; otters =otters.
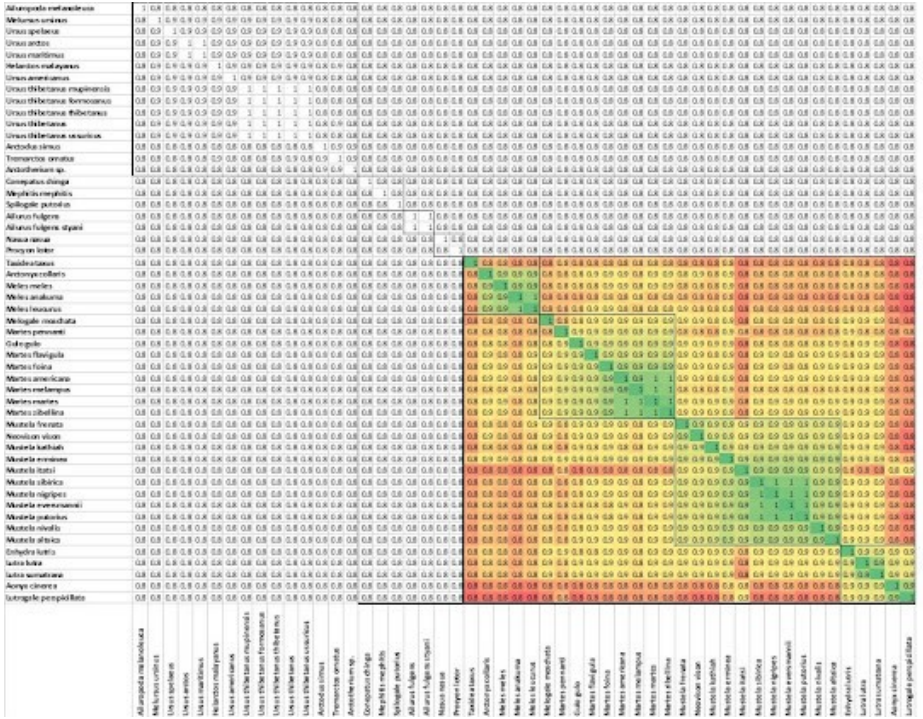

Figure 12 Same correlationmatrix as in figures 7, 8, 9 and 10, emphasizing differences within the family Mustelidae.

Would the otters stand a bit apart? After all, otters are biologically meaningfully different from the other species here - swimmers and piscivores.

Cserhati interprets figure 7 (his figure 3) as if the five known taxonomic families appear from the clustering.

> *In this heatmap we can see two main, large clusters as well as three smaller ones.*

There is no statistical underpinning of five clusters. Cserhati says five clusters, but then a silhouette plot would be expected that makes that clear. The silhouette plots given by Cserhati show something else.

The mtDNA part of the BMC Genomics article contains a curious stand-alone sentence among its description of the phylogenetic trees:

> S*upplementary figure 3 shows the average silhouette width according to the number of clusters, with an average silhouette width of 0.51 for two clusters*.
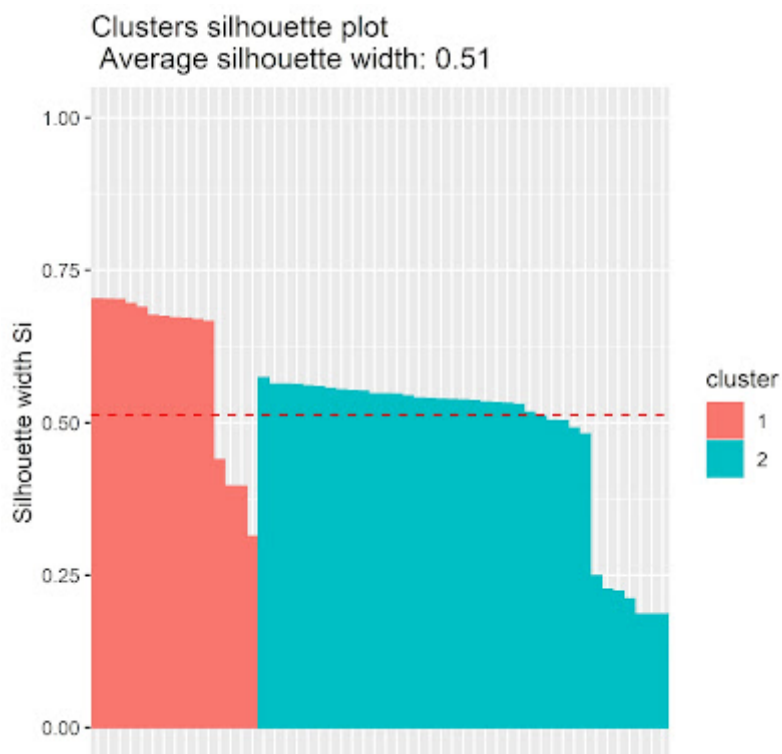

Figure 13. Supplementary figure 3 from the BMC Genomics artikel, silhouette plot with two clusters Left ursids, right Musteloidea. In this sihouette plot the 30 species of the marten family and the 7 species of Musteloidea that do not belong to the marten family can be recognized.

Cserhati has made silhouette plots for different numbers of clusters, as can be seen in Supplementary figure 3 of the CRSQ article.
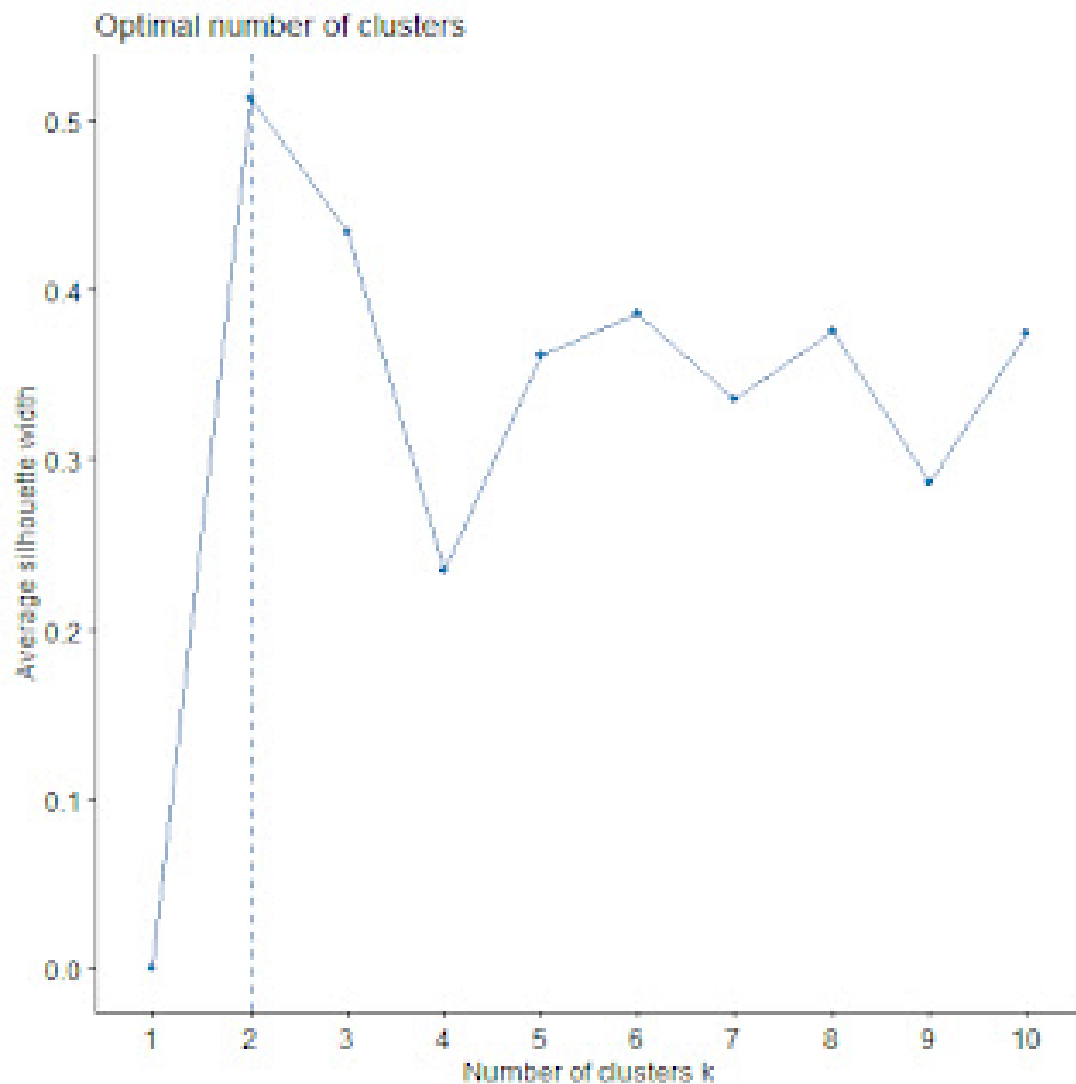
Figure 14. Supplementary figure 3 van het CRSQ artikel shows the average silhouette width according to the number of clusters, with an average silhouette width of 0.51 for two clusters as maximum. The optimal number of clusters is 2.

In fact, Cserhati finds two major clusters, namely the bear family Ursidae and the superfamily Musteloidea. He also indicates this in his table with the statistical data of the clusters, Table 3 of the BMC Genomics article.

Five clusters, the five classical families Ursidae bears, Mephitidae skunks, Ailuridae red panda, Procyonidae raccoons, Mustelidae mustelids is not better than eight clusters. Eight clusters with the four families Ursidae bears, Mephitidae skunks, Ailuridae red panda, Procyonidae raccoons and four subfamilies, Melinae badgers, Guloninae martens, Lutrinae otters and Mustelinae weasels within the marten family is also possible, as good as five clusters.

90

Amino acid sequence in the protein cytochrome b.

The gene for the protein cytochrome b is found on the mitochondrion. Cserhati decides to use the amino acid sequence of this protein for species clustering.

> *Cytochrome-b is a structurally conservative protein which does not mutate freely. It is often used to infer phylogenetic relationships between organisms (Meyer, 1994), and can so be used in baraminology studies as well.*

Cytochrome b was one of the first genes for which good techniques and good data became available. We are talking about popularity 30 years ago, before DNA sequences of the entire mitochondrial DNA became available. Since the, cytochrome b has fallen out of favour (Meyer1994). Genbank and the NCBI database have thousands of entries; finding data is not difficult.

Cserhati compares the amino acid sequence in 25 cats, 5 bears (but not the giant panda), the two subspecies of the red panda, the striped skunk *Mephitis mephitis* alone in its family, the coati *Nasua nasua* representing the raccoon family, and 17 species of mustelids . The heatmap that Cserhati gives of the correlation matrix for these species is given in the following figure.
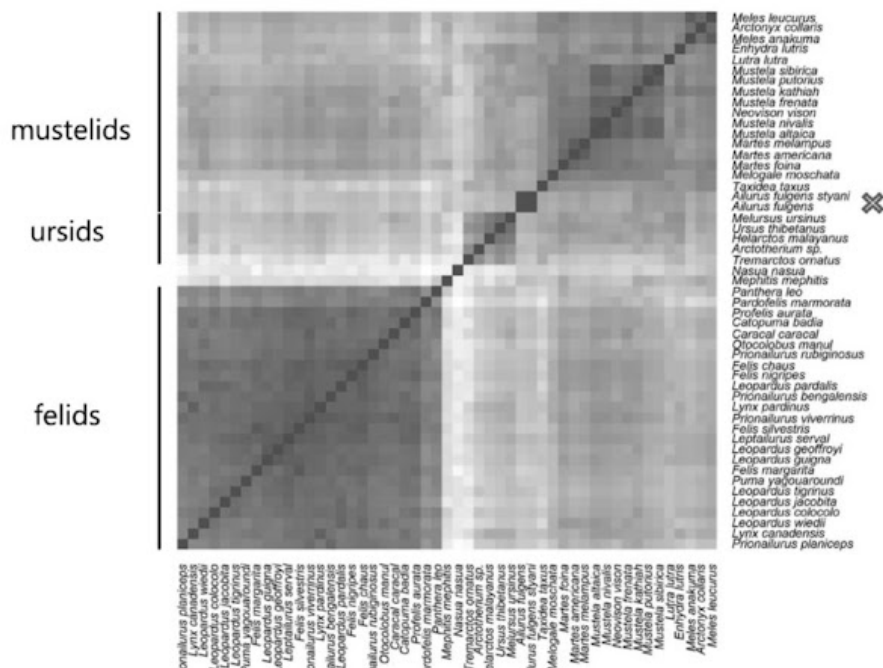

Figure 15: Heatmap indicating magnitude of correlations in cytochrome b amino acid sequence between species. Darker is more similar. The order of species on the axes is the same from top to bottom as from right to left; on the diagonal zw-no is the correlation with the species itself. The groups are indicated on the left. The two subspecies of the red panda are indicated on the right. The coati and the skunk are in this layout between the group of bears and the group of cats. Figure 4 CRSQ 2021.

Cserhati writes:

> *In the heatmap (Figure 4) we can see three main clusters and two species by themselves, N. nasua and M. mephitis. In the lower left there is a tight cluster of 25 felids. In the upper right there is a cluster of five ursid species, and another with 19 mustelid species.*

No silhouette plots are provided to substantiate three clusters, of cats, bears and mustelids + pandas, and two non-clustered species. A silhouette plot should have made clear why no cluster Musteloidea is reported. Given the heatmap, a cluster of Musteloidea is a possibility.

The red panda is without a doubt counted among the mustelids without any arguiing why. As can be seen from the heatmap (fig 15 here, fig 4 Cserhati) this is not obvious. The two subspecies of the red panda clearly form a block next to the species of the family Mustelidae; the correlations of the red panda with the mustelids are again lower than the correlations between the mustelids themselves, as can be seen from the gray values.

This becomes clearer in a red-green heatmap. The correlations can be found in Additonal File 3 of the article. I've changed the order of the families along the axes, and brightened up the correlation matrix for a clearer heatmap.
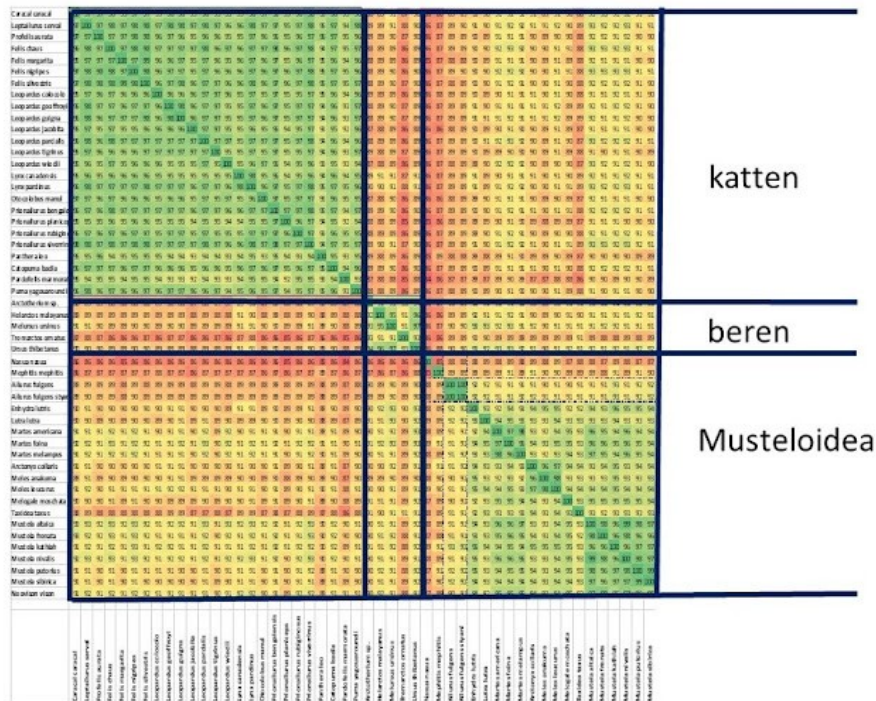

Figure 16. Heatmap indicating magnitude of correlations in cytochrome b amino acid sequence between species. Greener is more similar, redder is more different. Minimum Correlation: .84. The order of species on the axes is the same from top to bottom as from right to left; on the diagonal nw-zo is the correlation with the species itself. Figure 4 CRSQ 2021 with change of order. (katten = cats; beren = bears)

92

The cat group is homogeneous, the few bears are also homogeneous, and the superfamily Musteloidea mustelids, coati, skunks and red pandas appear to be quite heterogeneous. The coati is the topmost in the Musteloidea, followed by the skunk, and then the two subspecies of red pandas. coati and skunks differ Coloring only the superfamily Musteloidea.

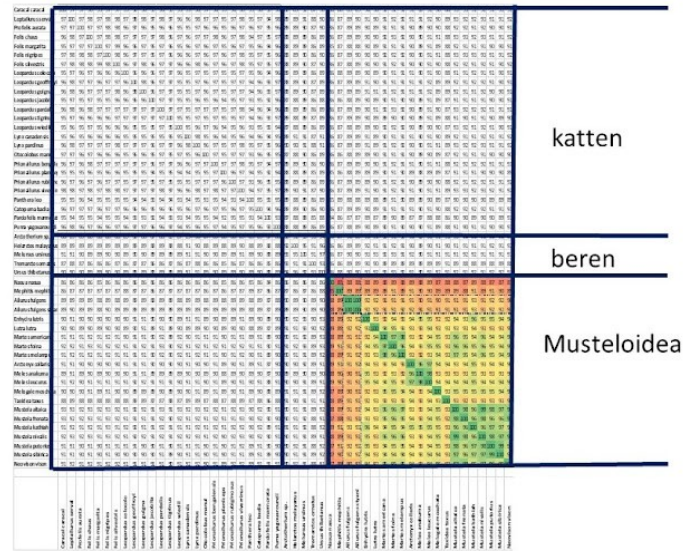

Figure 17. Figure 16, Musteloida colored green = similar red=different.(katten = cats; beren = bears)

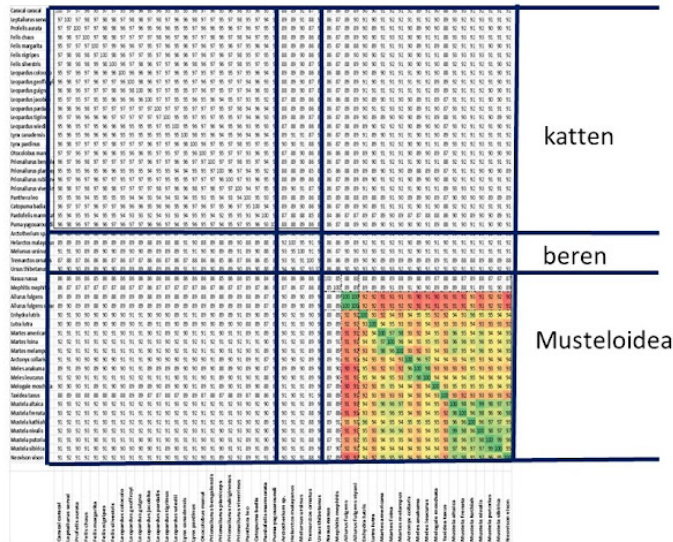Coloring only red panda and family Mustelidae:

Figure 18. Figure 16 red panda and Mustelidae colored green = similar red = different. (katten = cats; beren = bears)

The large difference of the coati *Nasua nasua* from the other species of the superfamily Musteloidea is the striking feature of the heatmap..

Cserhati uses a strange selection of species to determine whether the red panda belongs to the mustelids based on cytochrome b. Why 25 species of cats? Those cats don't matter much in the analysis; at most they drive the non-cats in a heat map to colors that are closer together. If we want to know where to place the red panda within the superfamily Musteloidea, we have to use species from the Musteloidea.

Cserhati does not give his Fasta file with amino acid sequences for cytochrome b, but he does give the numbers of the accessions he used. I looked up the amino acid sequence of cytochrome b for mustelids, bears, red panda, skunk, coati, and seven species of cats in the NCBI database. I supplemented this with the amino acid sequence of another skunk, and ten species of the raccoon family Procyonidae. This gives the following heatmap:
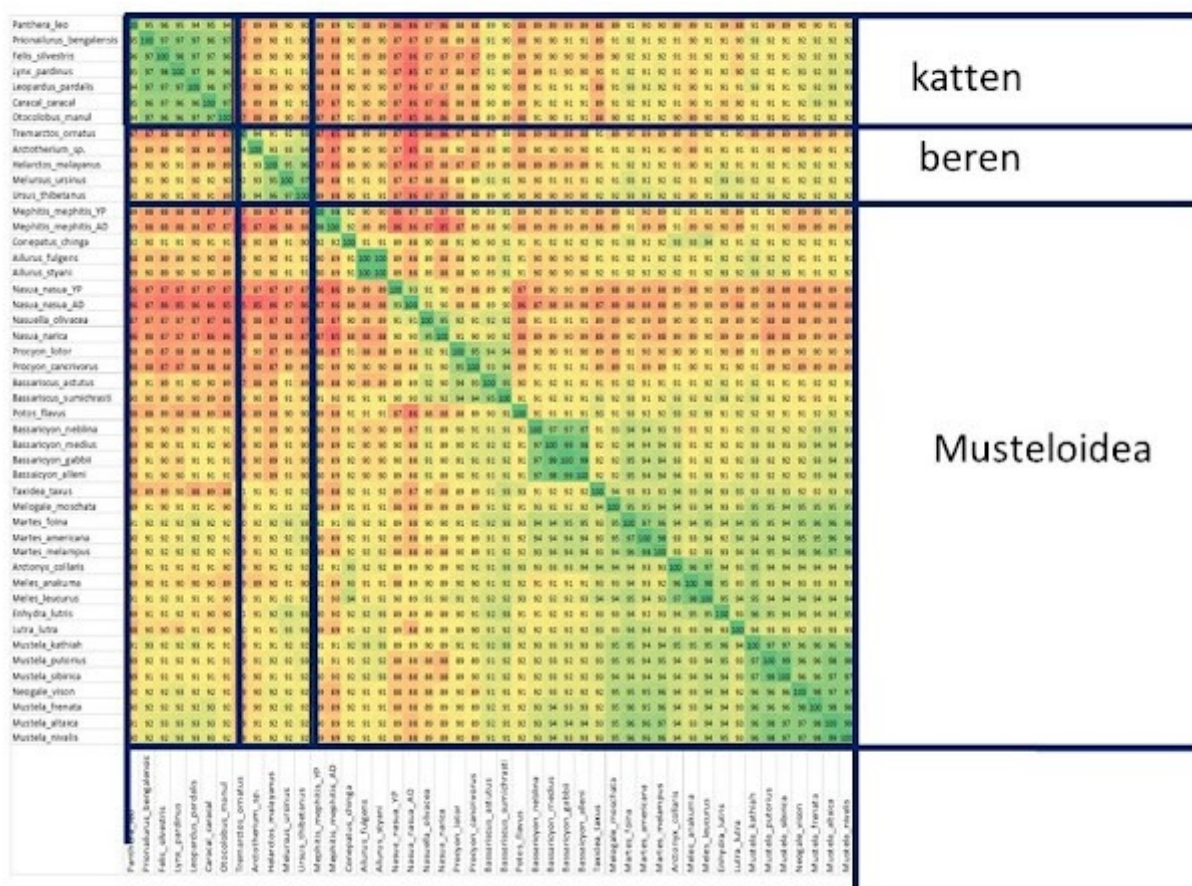

Figure 19. Correlationmatrix of cytochrome b amino acid sequences: fewer cats, more Musteloidea. Green = similar, red = different. Minimum correlation: .85. Katten = cats; beren = bears.

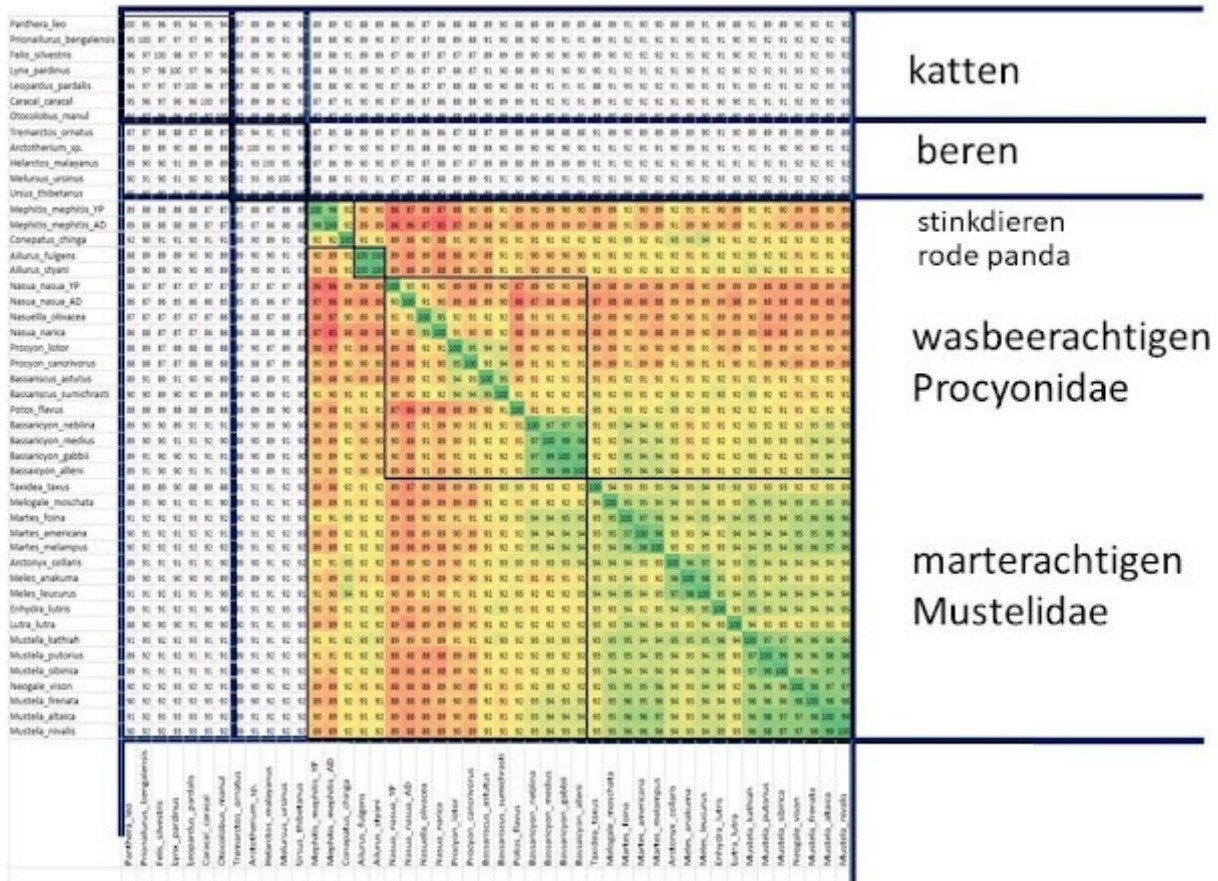The superfamily Musteloidea proves quite heterogeneous. In more detail:

Figure 20. Correlationmatrix of cytochrome b amino acid sequences: fewer cats, more Musteloidea. Green = similar, red = different. Minimum correlation: .85. Katten = cats; beren = bears. stinkdieren = skunks;

Within the family Procyonidae, some species are quite similar to the family Mustelidae, but other species are very different from the Mustelidae. One of the species that differs greatly is the coati *Nasua nasua*. The choice of only the coati in the Cserhati heat map makes the red panda more closely resemble the mustelids than is warranted across the whole of the Procyonidae.

Conclusions

The WGKS clustering gives three groups: the cat family Felidae, the bear family Ursidae and the superfamily Musteloidea.

The mtDNA clustering gives two groups: the bear family and the superfamily Musteloidea.

The cytochrome b clustering gives a cluster cats, a cluster bears and is unclear on the Musteloidea.

The heatmaps indicate that the red panda does not fall within the family Mustelidae in any of these three analyses.

It can also be seen that clustering does not accurately represent biology. The mtDNA analysis gives two clusters, but the heatmap clearly shows how much detail is lost during clustering.

Sloppiness:

Reporting on the mtDNA heatmap Cserhati writes in BMC Genomics *"Three larger clusters and two smaller clusters are visible in the heat map."* (pg 4-5, flanking figure 1; reporting on the same data in CRSQ, Cserhati writes "*In this heatmap we can see two main, large clusters as well as three smaller ones.",* pg 80

Page 78: whole genome sequence for Procyonidae is mentioned as downloaded

Page 79, Table 1: *Spilogale gracilis* in cluster 3, the cats

Page 81: "... the WGKS analysis, where the average silhouette width was 0.8 for classifying *M. mephitis* ..." It is *Spilogale gracilis* in the WGKS analysis, not *Mephitis mephitis.*

In Figure 2, the text is not in accordance with the figure. The figure should have given 'musteloids' on the left, not 'mustelids'.

 ***

Cserhati, M., 2021, Classification of the Enigmatic Red Panda (Ailurus fulgens) Based on Molecular Baraminology-Based Analysis, Creation Research Society Quarterly 58 (2): 76-84

M.T.N Tsuchiya, R.B Dikow, K.P Koepfli, P.B Frandsen, L.L Rockwood, J.E Maldonado, 2021. Whole-Genome Sequencing of Procyonids Reveals Distinct Demographic Histories in Kinkajou (Potos flavus) and Northern Raccoon (Procyon lotor) Genome Biology and Evolution, Volume 13, Issue 1, January 2021, evaa255, Published online: 17 December 2020

Hennigan, T. 2010. The case for holobaraminic status in bears (family Ursidae) and the implications within a creation model of ecology. CRSQ 46(4):271–283.

Meyer, A. 1994. Shortcomings of the cytochrome b gene as a molecular marker. Trends in Ecology & Evolution 9(8):278–280