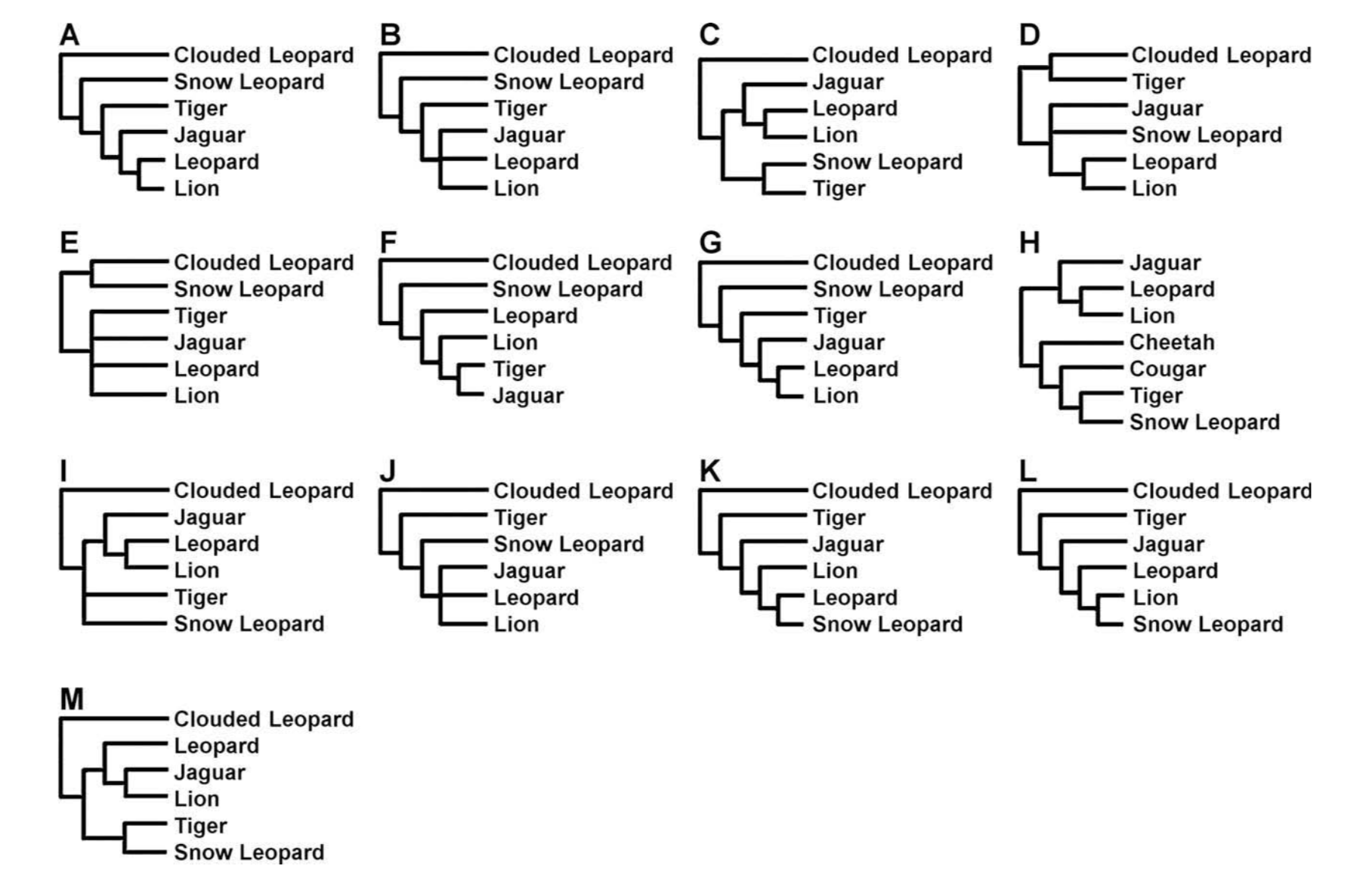
# Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, Panthera (Carnivora: Felidae): Review and Tree Reconstruction

Alexandra González

Phylogenetic Relationships of Big Cats; Genus Panthera

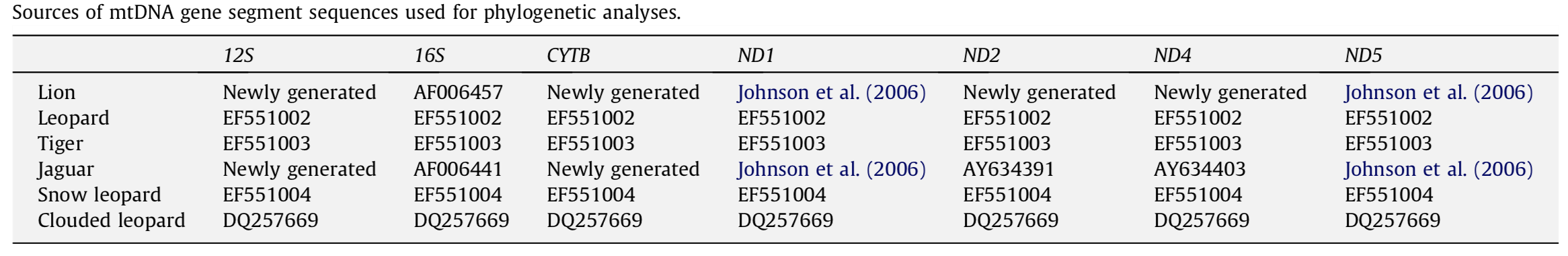
Based on Brian W. Davis, Gang Li, William J. Murphy article, *Supermatrix and species tree methods resolve phylogenetic relationships within the big cats, Panthera (Carnivora: Felidae)*, I will be working to form the phylogenetic tree results that they got when they used three methods, in which resulted in three topology phylogenetic trees. The species that they used were the lion (Panthera Leo), Leopard (Panthera Pardus), Tiger (Panthera Tigris), Jaguar (Panthera Onca), and Snow Leopard (Panthera Uncia). There is also an outgroup used, the Clouded Leopard (Neofelis Nebulosa). The problem that they expressed was that the evolutionary history of the five big cats has been concealed because there is very little information provided in the fossil record, as well as rapid radiation that took place during the Pliocene era (~5.3-2.6mya), speciation between the genus Panthera less than 1mya, and possible hybridization between these lineages after their separation as individual species.



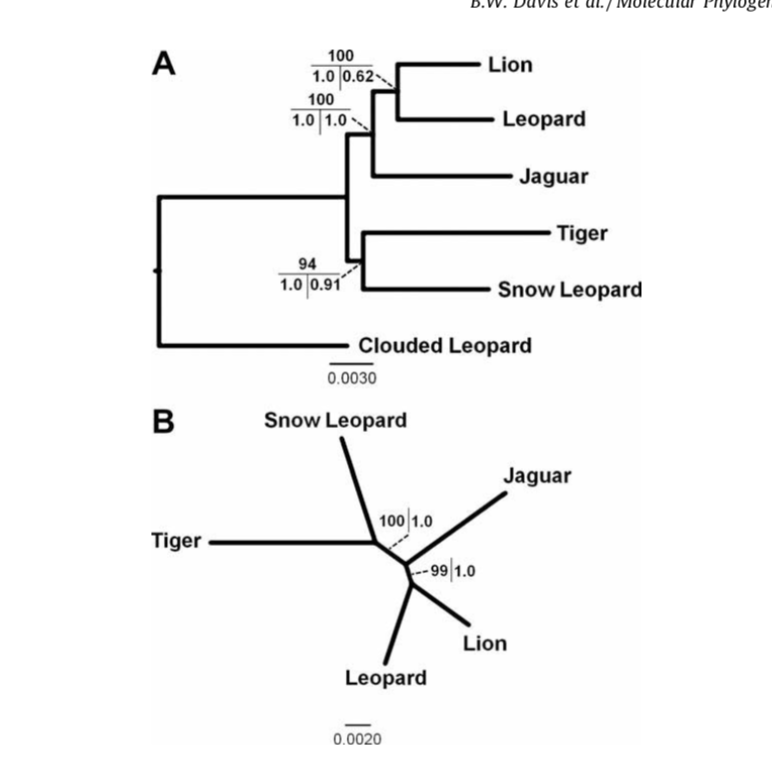
Various hypothesis and theories have been formed to group these lineages together and create the “correct” tree based on morphological or biochemical data. However, lack in phylogenetic analysis has led to many misidentifications between species relatedness.

The purpose for this study was to find a way to change how this lineage contain vulnerable and endangered species like the tiger and jaguar due to human influence in the ecosystem. Conservationists wished to use genetic data to make an action plan that could help prevent these numbers going down even more and create a balance in the ecosystem. They were conducting various possible trees to link them on from the other by using DNA sequence data and GenBank. In order to get their date, they used GenBank and Blast to use multiple genomes of all the species found in the genus Panthera.

To obtain gene sequences and mtDNA sequences, researchers used GenBank and BLAST. Some sequences and accession numbers were taken from other studies such as Johnson’s W.E. journals. In their DNA sequencing data, they used sex chromosomes and mitochondrial genomes, which included 39 single-copy of Y chromosomes, 4 mitochondrial and 4 autosomal gene segments. The methods that they used to get the phylogenetic trees were Maximum Likelihood, Bayesian Phylogenetic Interference, and Bayesian Estimation of Species Trees (BEST).The purpose of using these methods was to make a supermatrix of these species in order to determine whether any of these species had a monophyly with each other based on their similar topologies They used ClustalX 2.0.3 in order to align the sequence they needed. The researchers used PAUP to perform Maximum Likelihood tree searchers and MrBayes analysis to show summary of tree statistics.

The table above contains all the mtDNA sequence used in the study.

When the researchers run the Bayesian Interference, they ran 1.5 mil generations and saved every 100th tree, as well as discarding the first 250000 as burn-in for the individual gene segments of the 6 species. For the 6 matrices, the MCMC algorithm ran for 3.o mil gen., with every 100th tree saved and the first 750000 discarded. BEST analysis was then used to construct a species tree from individual Bayesian genes and its MCMC algorithm ran for 10 mil gen., saving every 1000th gene tree and discarding the first 1,000,000 as burn-in.



Results showed that the Clouded Leopard was established as an outgroup in the rooted tree after Maximum Likelihood topology was read. Two distinct clades were formed as well, one where the lion, leopard and jaguar share a specific set of common characters and separated them from the tiger and snow leopard. The unrooted tree showed the bootstrap results (left) and the Bayesian Posterior Probabilities (right).

-*Figure 6*

The ML analysis results based on the complete Supermatrix showed a Lion-Leopard monophyly, with the jaguar being a sister taxon, and Tiger-Snow Leopard monophyly. The X chromosome partition also recapitulated lion–jaguar monophyly with significantly greater support using BI than ML, consistent with the non-conservative nature of the Bayesian method. However, little support was seen for the tiger and snow leopard to show a distinct monophyly; the rooted mtDNA partition did not support tiger–snow leopard monophyly, and only received low support when unrooted. The branch length between the tiger and snow leopard in the unrooted tree (shown above) is far off from each other.

Overall results of the study included that there was an evident monophyletic relationship between the Lion and Leopard. The article also that two distinct clades were observed within the Panthera species: Lion-Leopard clade with the Jaguar being part as a sister taxon, and the sister relationship Tiger-Snow leopard clade, which showed low support for monophyly during bootstrapping the unrooted tree (Fig. 6). Their results gave them a new light in the speciation genomics and understanding the historical events that helped portray the origin of this lineage.

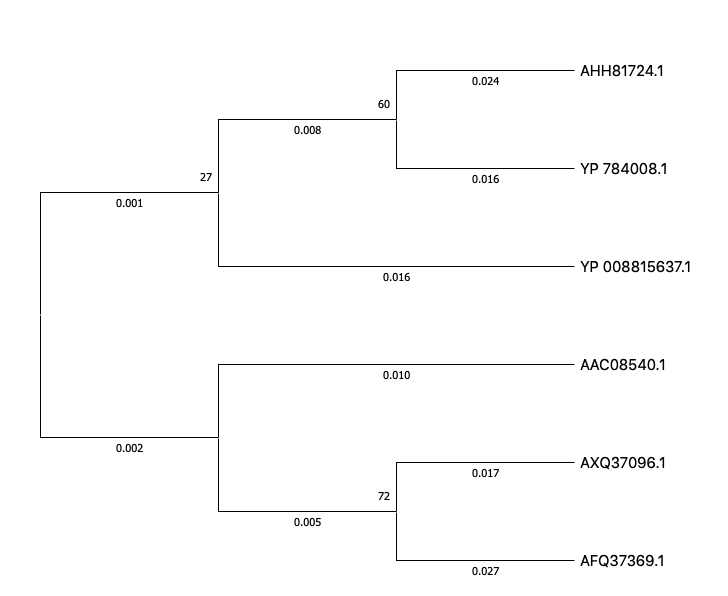
My plan was to re-make their phylogenetic tree of having the lion and leopard as a monophyly, while the jaguar is a sister taxon for the group, as well as sister taxon for the leopard and tiger. First thing to do was to find the data set they used and used the Maximum likelihood and the Bayesian Phylogenetic Interference methods to see if I have conducted topological phylogenetic trees. The sequence that I used in order to attempt to reconstruct this tree were mitochondrial whole genomes of the five big cat species from Blast.

|  |  |
| --- | --- |
| **Name** | **Accession Number** |
| Panthera Leo | AHH81724.1 |
| Panthera Onca | YP\_008815637.1 |
| Panthera Tigris | AAC08540.1 |
| Panthera Pardus | AXQ37096.1 |
| Panthera Uncia | AFQ37369.1 |
| Neofelis Nebulosa | YP\_784008.1 |

Once these sequences are downloaded, I made a new directory with these sequences in both FASTA and Nexuss formats.

**Results:**

Used the cytochrome B protein sequence and aligned them with MPI Bioinformatics Toolkit. Afterwards, I built a ML and Max Parsimony tree through MEGAX program with a bootstrap replication of 100.



This tree depicts a Max Likelihood method, run by MEGAX and the trees for the heuristic search were obtained with the Neighbor-Joining method to a matrix of pairwise distances estimated with the JTT model. The branch lengths are shown for each lineage.

This tree does not depict a monophyly between the leopard and lion. It depicts a close relationship between the lion and clouded leopard, while the leopard depicts a close relationship with the snow leopard. This error could be caused by using morphological similarities instead of phylogeny between lineages when evaluating ML.

While looking for the same sequences for the cytochrome b whole genome used in the paper, difficulty in finding the correct alignments led to me using different sequences of the six species of Panthera from the study in order to be able to conduct a Maximum Likelihood test. The tree above does predict that there was a relationship with the jaguar and the lion. However, the addition of the clouded leopard sequence in this case seemed to closely relate to the lion’s protein sequence.

The tree on the left depicts ML analysis without the Clouded Leopard sequence. The lion and jaguar are grouped as sister taxa. The snow leopard and leopard are grouped as sister tax as well, with the tiger closely related to both of them.

A screenshot of a cell phone

Description automatically generated

**Overall Progress:**

After many ups and downs with finding different sequences in order to have a successful alignment in order to conduct a Maximum Likelihood analysis with bootstrap. Trying to conduct MrBayes with these sequences was very difficult due to my very limited knowledge on how to wrong the program, as well as trying to use just one type of gene sequence proved to not have the same results than the journal article had. The article used multiple gene sequences for each species and conducted various tests in order to get the most significant tree to prove the monophyly between the lion and leopard. My analysis did not yield the same results because I was not able to use all sequences. In the future I wish to expand my knowledge and skill on how to properly align various mtDNA or protein sequences, as well as be able to run MrBayes in order to attempt to build the correct tree.