ANALYZING THE TRANS-HEMISPHERAL MOVEMENT OF *ACIPENSER* SPECIES BY THEIR mtDNA PHYLOGENETIC RELATIONSHIPS

Catherine Gohar, Anirudh Prasad, Madhav Rajkondawar, and Luis Cruz Reyes

ABSTRACT: Taxonomic naming of organisms leads to species which are connected by their evolutionary ancestors. The family of Acipenseridae has existed for over 200 million years through the delicate evolution of descendants. Within this time, the genus of species within it has diversified and spread across the world. Several years ago, however, groups of sturgeons became threatened while others remained relatively stable. This study focuses on the percent of genetic similarity in mitochondrial coding genes for the Acipenser genus of sturgeons, between the eastern and western hemispheres to ask why some sturgeons might be threatened. We hypothesized that our Principal Coordinate of Analysis will show a genetic crossing of the Acipenser species through the mixing of western and eastern species in one cluster. We imported FASTA files for the mitochondrial DNA (mtDNA) of all 17 Acipenser species and sorted the DNA sequences into different protein CoDing Sequences (CDS) in a Python program. By utilizing Basic Local Alignment Search Tool (BLAST), the Python program compared two species at a time to get the percent similarities. With this data, we were able to make a multidimensional scale (MDS) of species similarity using a dissimilarity symmetric matrix. This matrix was then used to create a Principal Coordinate Analysis (PCoA) in XLSTAT. The PCoA plot helped us learn that the species were instead separated into two clear clusters, one of only eastern species and another of both western and eastern species. Our alternative hypothesis was supported, and we note that sturgeons of the Eastern hemisphere, especially the Atlantic coast, appear to crossbreed less often. This might have resulted in lower genetic diversity and correlates with eastern-hemisphere sturgeons being threatened by human and environmental causes.

Keywords: Acipenseridae, Acipenser, Principal Coordinate Analysis (PCoA), Sturgeon, Genetic Similarity, Dissimilarity Matrix, Genetic Analysis, Mitochondrial DNA (mtDNA), Phylogeny, Python, Basic Local Alignment Search Tool (BLAST)

Introduction

Over the past few millions of years of life, most animals have evolved drastically from their ancestors. Sturgeons have remained relatively like their older ancestors despite many environmental changes and migration events (Rajkov et al. 2014). There are about 27 species given the common name “sturgeon”, all of which fall under the Acipenseridae family in taxonomic naming. While they have evolved without much divergence from their ancestors, recently sturgeons have become threatened or critically endangered due to ocean pollution or competition by human and other causes (Chassaing, Olivier, et al. 2016). We would like to focus on the phylogeny of these species, so that future research can focus on the genetic loss or fixation of the sturgeon gene pool, which may lead to their downfall. This project studies the Acipenser genus, holding 17 of the common sturgeon species spread throughout the Northern hemisphere (Rajkov et al. 2014). Even with a fossil history containing two of the 17 extant species, the general phylogeny of Acipenser is up for debate because of complicated genetics

and limited specimens to experiment with (Bemis et al. 1997). There is one common phylogenetic tree, but small node switches, like A. dabryanus and A. transmontanus's placement in comparison to the A. stellatus sturgeon, are still questioned (Rastorgeuv et al. 2008). Through a mitochondrial DNA (mtDNA) analysis of the 17 Acipenser sturgeons, we also hope to elaborate on the relative genetic distances of sturgeons.

There has been some controversy in using mtDNA for constructing phylogenies as they do not always match the phylogenies made for nuclear DNA similarities. However, the strength in mtDNA that made us choose it over nuclear DNA was that there is little recombination of mtDNA overtime (David 2005). Since we are also studying at the genus-species level, we can safely avoid homoplasy (where a new trait for a gene arises twice, separate from their common ancestor) that sometimes confuses phylogenies. This allows for more confidence in our PCoA results, assuring us that the similarities are indeed due to common ancestors and not convergent evolution. In our previous experiment comparing sturgeon's entire mtDNA genomes, we noticed that western and eastern hemisphere sturgeons were more like each other than they were to individual hemispheres. Wanting to expand on that, we are now comparing the protein-coding genes of Acipenser mtDNA, calculating gene similarity for every combination of sturgeon, and producing a Principal Coordinate Analysis (PCoA) based on their genetic distance for dissimilarity.

For the Acipenser mtDNA genome, there are 13 protein CoDing Sequences (CDS) which may differ significantly across sturgeon species (Mitofish 2020). The mtDNA of Acipenser species also has 22 tRNA sequences, and 2 rRNA sequences, but we will not be comparing those genes in this study. By comparing similarity (or dissimilarity) between specific protein-coding genes that may have changed over time, we can create a diagram showing how closely related each species is. Our hypothesis is that, as cross-hemispheric sturgeons are more likely to be related, their genetic distances will be closer and represent one rather than two clusters. This should also suggest that sturgeons have crossed often in their history instead of staying in their relative hemispheres. Alternatively, there is a possibility that sturgeons will be clustered based on their hemispherical location, with western and eastern species being separated into two clusters on the PCoA.

While we could run genome-annotating programs ourselves and sort through the hundreds of combinations of CDS strings per sturgeon, computational biology has made it so processes that take days can be completed in a maximum of five or ten minutes. Our group chose to code a Python program which would splice protein coding genes in each sturgeon genome for us, compare our data in a Basic Local Alignment Search Tool (BLAST), and produce a PCoA displaying their evolutionary distance. Nucleotide BLAST is a handy tool which calculates percent similarity between two strands of nucleotides, while accounting for minor base-pair changes that might normally offset two sequences (BLAST 2020). Python is a programming language that has been a massive help to DNA sequencing and sorting as a high-level programming language (Python 2021). Bioinformatics, a field devoted to programming tools for understanding experimental results in biology, has exploded in the past few decades because of these advantages in saving time and money. With these tools at our reach, we were able to complete the bulk of the project with relative ease.

Methods

First, we imported all 17 Acipenser mtDNA genomes into Python via FASTA files, along with the 17 annotated Acipenser files which held the CDS start and stop locations. These locations are defined as the number of nucleotides down the mtDNA where each gene begins and ends. These were acquired through MitoFish's annotator

program, which recognizes the start and stop of each CDS (Mitofish 2020). After each of the start and stop locations were made accessible through Python, we called a method to pull each CDS string directly from the Acipenser FASTA files and saved them into a global variable array. This was so they could be used to compare species with nucleotide BLAST.

We used the start and stop indexes given in the annotation files to save each CDS from the Acipenser FASTA files, which totals to about 13 sequences. Each of these start/stop indexes corresponded to one Acipenser FASTA file. We did this by opening the annotation file to parse, going through each line, and only reading the lines that have the string "CDS" in them (avoiding tRNA and rRNA strings). After we obtained the CDS lines from the annotation, we stored the start/stop integers as indexes. The final step was to open the FASTA file of the species and get the substring using the start/stop indexes, then store the CDS string into the array that holds all the other CDS values. We ran a for-loop on the array to individually collect information such as the start/stop indexes for each CDS before moving to the next sequence in the list. This saves room in the code by repeating the same actions for different items in the array, until each CDS has been accounted for. We printed the array at the end of the for-loop going over all CDS lines, to make sure all sequences are stored in the array correctly.

To compare two of the Acipenser species' CDS, we used the two global variable arrays created by the previous method to save all CDS values. We also invoked BLAST through the Python code we wrote to obtain the similarities of these sequences. First, we created two temporary FASTA files to store the CDS for the two species to be compared. Next, we included the environment variable to make sure that BLAST runs on Windows/Linux. Then, we invoked "makeblastdb" to run all the commands through the .exe file using BLAST. This way, we can run every program through Python instead of having to use the command prompt. We also have a command that takes a "query" and "subject", which are the two temporary FASTA files that save the CDS. This command is sent to the system to get a .txt output for the similarity percentage between the two species. We finally close the two files, so that the content can be erased and replaced by the new CDS when the loop starts again through the global variable arrays. A limitation to our current project is due to BLAST for only printing 12 percentages, while 13 should have been seen. All the files consistently had 12 percentages, so we continued with 12 percentages for the results file. In the future, we will fix our code to understand why exactly one gene comparison was missing. To work around this, we took an average of all these CDS percentages for every species and saved them to a .csv file along with the names of the species compared. The .csv file represents a collection of all the species' average similarity for all their CDS. We ran the program multiple times to test that the .csv file, saved all the correct names of each species, and compared their respective percentages.

Lastly, we chose one species to get the resulting percentages for all possible unique combinations of genetic similarity. We then compared it to every other species in the entire list and ran it until we made every possible comparison. These average percentages are stored in the Supplementary Data 1 file. We simplified the number of combinations to unique pairs by removing the duplicate percentages using the Microsoft Excel function "remove duplicates". Then, we used these results to create a multidimensional scale (MDS) of species similarity. We converted our data to a dissimilarity symmetric matrix by reversing the percent similarity to dissimilarity and arranged the matrix as seen in the Supplementary Data 2 file. A similarity (or dissimilarity matrix) is a table of all species on both the x and y axis, with their respective percentage of mtDNA genome similarity in each cell that lines up with a species pair. If it is a similarity matrix,

a value of 1 is 100% mtDNA similarity, while a value of 0 means 0% mtDNA similarity. In a dissimilarity matrix, a value of 1 means 0% mtDNA similarity, while a value of 0 means 100% mtDNA similarity. We needed to create a dissimilarity matrix (opposite of what BLAST was outputting) to correctly format our MDS. This dissimilarity matrix helped produce the Principle Coordinate Analysis (PCoA), which is saved in the Supplementary Data 3 file. After running our Python program and comparing each species, we compiled our obtained percent similarity results in Table 1. Then, we created Table 2 by subtracting the percent similarities from 100 to create a difference, or "dissimilarity", table. Using this difference table, we ran the differences through a PCoA function for XLSTAT to obtain the MDS, or PCoA, graph seen in Table 1. From this graph, we analyzed whether the Acipenser species crossed hemispheres once (with two evolutionary distance clusters) or multiple times (with one evolutionary distance cluster).

Results

We found that for the most part, Eastern Acipenser species clustered together, which was represented as Cluster 1 in Figure 1. As can be seen in Cluster 2 in Figure 1, Eastern Acipenser species Mikadoi, Dabyranus, Schrenckii, and Dabryanus x Schrenckii loosely clustered together with Western species Medirostris and Transmontanus. Interestingly, the Western species Acipenser Brevirostrum clustered tightly with the Eastern species Cluster 1 on the left side of the graph. Furthermore, the hybrid species A. Schrenckii x Dauricus is an outlier, distributed to the bottom right side of the plot. The Western species A. Oxyrinchus and the Eastern species A. Sturio were outliers while still being grouped together, shown in Figure 1. Overall, we can see that our alternative hypothesis was supported by a clear separation of two species groups, specified as Cluster 1 and Cluster 2 (Figure 1).

Analysis and Brief Conclusions

Upon analysis of our results depicted in Figure 1, we can glean that the Eastern species were the basal descendants from a phylogenetic perspective due to the tight distribution of Cluster 1. Since Acipenser originated in the Eastern hemisphere, we can gather from the data that the Eastern species in Cluster 1 did not cross hemispheres as much, if at all, in comparison to the species located in Cluster 2. However, due to the tight nature of the distribution, we can assume that the Eastern species of Cluster 1 interbred frequently within their hemisphere to the point that their genetic makeup became very similar to one another. The presence of A. Brevirostrum in Cluster 1 also indicates a limited trans-Atlantic crossing-over event.

Cluster 2 has a looser and more "mixed" nature of the distribution, in comparison to Cluster 1. This leads us to believe that cross-continental breeding was far more prevalent in the Eastern and Western species belonging to this cluster. This is supported by a simple analysis of the geographic locations of these species. The Eastern species of Cluster 2 are in Russian, Chinese, and Japanese waters, and the Western species of Cluster 2 are located in either the Pacific Ocean or the Western seaboard of the United States and Canada. This is especially true of the Western species A. Transmontanus, which resides on the Pacific Coast near California and even as high up as Alaska (Dershimer). Although they are from different hemispheres and reside on different coasts, they do share the same ocean.

This relationship between Eastern and Western species in Cluster 2 provides us with evidence to conditionally accept our hypothesis that some Acipenser species crossed over often, particularly those near the Pacific Coast (meaning that one cluster is probable). While there are certain cases of crossing-over with Acipensers near the Atlantic Coast, those crossing-over events are too few to reliably

accept our hypothesis, and we opt instead to conditionally accept the alternative hypothesis regarding species along the Atlantic Coast.

Discussion

After getting the results from the PCoA, we noticed that there were some Acipenser species that had significant dissimilarities, one of them being Acipenser transmontanus and Acipenser baerii, which were on two opposite sides of the PCoA plot and in separate clusters, as seen in Figure 1. In Fiske et. al's (2019) study of determining ploidy in white sturgeon, the study compared entire erythrocyte long-axis lengths obtained using Coulter counter and blood smears to find evolution over time. The flow cytometry analysis from this experiment showed significant genomic differences in the Acipenser baerii and Acipenser transmontanus species (Fiske et al., 2019). This was found to be true in our PCoA (Figure 1), where we also noticed drastic dissimilarities between the two species compared. Subsequently, this shows that the results we found make sense within the context of this study, where Acipenser baerii and Acipenser transmontanus are found in separate clusters and found to be significantly different from one another.

Another major pattern we noticed in species similarity was that many of the Eastern Acipenser species were clustered together, as we see in Cluster 1 in Figure 1. We also notice that in Cluster 2 in Figure 1, the Eastern Acipenser species mikadoi, dabyranus, schrenckii, and dabryanus x schrenckii have a weak and loose clustering with the Western species medirostris and transmontanus. As for the Western species A. brevirostrum located in Cluster 1, we can hypothesize that while it is a Western species, it did not cross over or interbreed as many times in its genetic history as the other Western species, thus maintaining a similar genetic makeup with the other Eastern species of Cluster 1. This is supported by the Bayesian Inference (BI) phylogenetic tree created by Krieger et. al. (2000) during their study on Acipenseriformes' phylogenetic relationships. In their BI tree, A. brevirostrum shares a common ancestor with many of the Eastern species in our study, while the Western species in our study are grouped further down, farther along the tree than A. brevirostrum. Another interesting comparison between our studies is that A. oxyrinchus (Western species) and A. sturio (Eastern species) are depicted in a clade in the BI tree, which explains why these two species are grouped together in Figure 1 of our study.

In Brown et al's (1996) study of length variation and sequence divergence in the mtDNA of four Acipenser sturgeon, the study compared particular protein spots in a species' entire DNA (Li et al., 2010). The results of this research found A. transmontanus and A. medirostris, which are on the Pacific coast, were more closely related than A. fulvescens and A. oxyrinchus, which are the Eastern species (Brown et al., 1996). Our results in the Principal Coordinate Analysis showed identical results for the A. transmontanus and A. medirostris being closer related than A. oxyrinchus, which was drastically isolated on the graph. The two separate Clusters in Figure 1 also illustrated significant differences between the two hemispheres and demonstrated crossing multiple times, as shown in Brown's study.

Having an in-depth analysis of the two literature papers mentioned above helped our project in facilitating the analysis of the results we found. Comparing and checking our results confirms our rationale behind them and gives us confidence as to why we got this data from our research on the Acipenser species.

Aside from migration patterns of sturgeons, genetic diversity has been a factor that affects their fitness in different environments (Nelson & McAdam 2012). Our own data revealed Acipenser sturio being isolated by almost all other sturgeons besides A. oxyrinchus, and this sturgeon is known to have less similarity

(as seen by their mtDNA distance in Figure 1) compared to other sturgeons (Chassaing et al. 2016). They also have competition with A. oxyrinchus and have been struggling to spread to other estuaries due to a small genetic pool to evolve from. A similar genetic situation is seen within Michigan sturgeons or A. transmontanus. They often stay within their own lakes, revealing dissimilarities within their own species' DNA (Nelson & McAdam 2012). Fewer migrations may contribute to the declining and endangered sturgeon populations, starting with A. sturio and A. oxyrinchus in Europe. While not apparent in the stable population of A. transmonatnus, they are currently in a similar situation, being isolated in the different Michigan lakes, and may follow similar patterns with an unforeseen disaster.

Based on our findings, it may be prudent to research why species along the Atlantic Coast are less prone to crossing-over between hemispheres than species along the Pacific

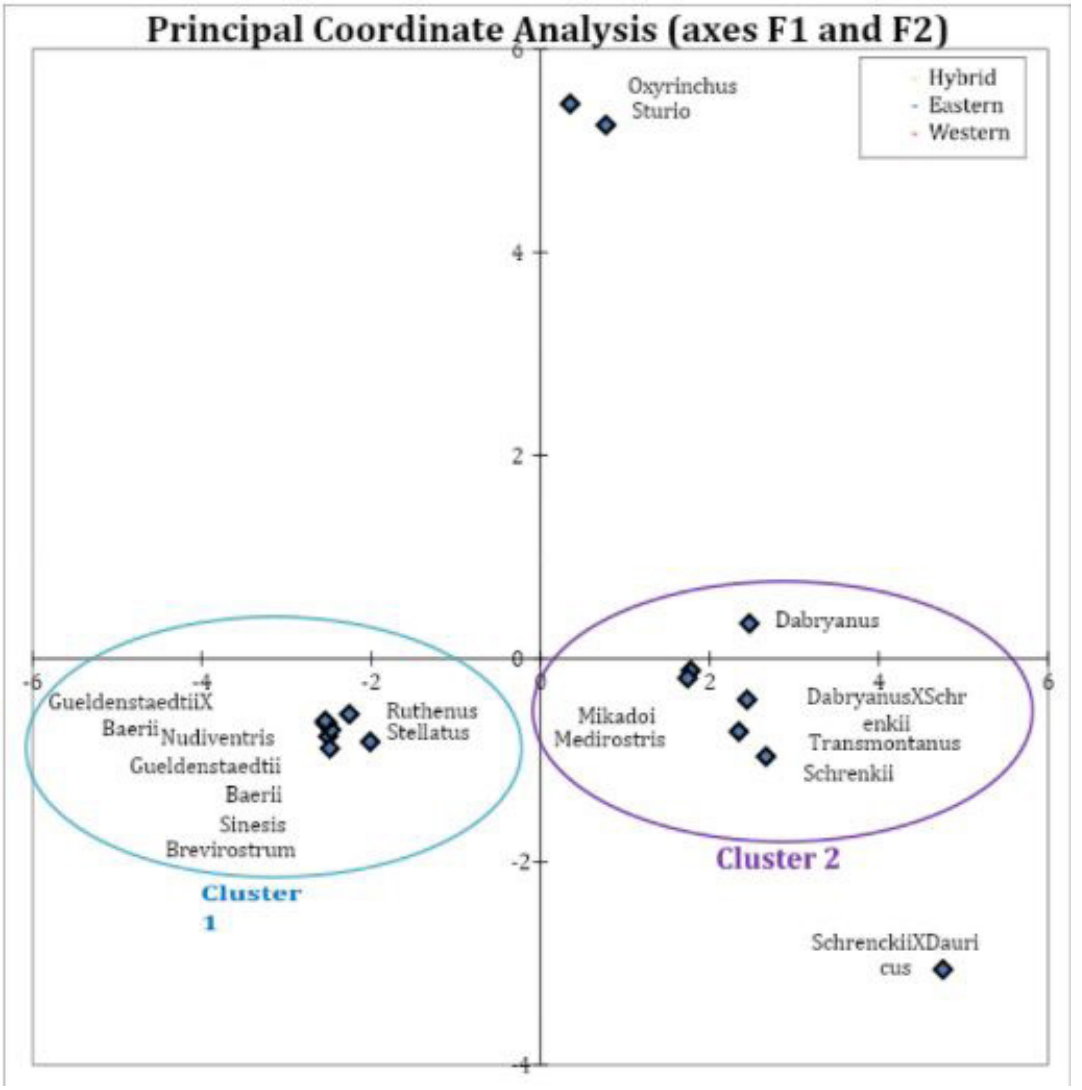

Figure 1: PCoA of all Acipenser species comparisons. The y and x axis show the distance of dissimilarity between species. Closer clustering represents little dissimilarity (or higher similarity), while long distances between species mean they have high dissimilarity (or little similarity). Note that blue species are Eastern, red are Western, and orange is the hybrid cross of Huso dauricus and Acipenser schrenckii which are also part of the Acipenser genus.

Coast. There is a noticeable difference of observed crossovers between species along the Atlantic and Pacific Coasts. It would be worth researching reasons why species along the Atlantic Coast are less prone to cross over than species along the Atlantic Coast, to preserve endangered Acipenser sturgeons more efficiently.

Acknowledgments

This project would not be possible without Jesse Zaneveld, who listened to both the unnecessary and groundbreaking questions we asked, as well as made time to review our programming and phylogenetic analysis. Thanks to the many students in Bioinformatics 383, who gave feedback and readily explained field-specific jargon, software, or science.

References

Bemis, W. E., Findeis, E. K., Grande, L. "An overview of Acipenseriformes". *Environmental Biology of Fishes*, vol. 48, 1997, pp. 25–71, doi:10.1023/A:1007370213924.

BLAST: Basic Local Alignment Search Tool. U.S. National library of Medicine, *National Center for Biotechnology Information*, 2021, https://blast.ncbi.nlm.nih.gov/Blast.cgi. Accessed 9 Jun. 2020.

Brown, James R., et al. "Length Variation, Heteroplasmy and Sequence Divergence in the Mitochondrial DNA of Four Species of Sturgeon (Acipenser)." *Genetics*, vol. 142, no. 2, 1996, pp. 525–35, https://pubmed-ncbi-nlm-nih-gov.offcampus.lib.washington.edu/8852850/?from_term=DNA+comparison+acipenser+species&from_pos=3

Chassaing, Olivier, et al. "Phylogeography of the European Sturgeon (Acipenser Sturio): A Critically Endangered Species." *Molecular Phylogenetics and Evolution*, vol. 94, 2016, pp. 346–357, https://doi.org/10.1016/j.ympev.2015.09.020.

Daniel Rubinoff, Brenden S. Holland, Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference, *Systematic Biology*, vol. 54, no. 6, 2005, pp. 952–961, https://doi.org/10.1080/10635150500234674.

Dershimer, Charles. "Acipenser Transmontanus (Oregon Sturgeon)." *Animal Diversity Web*, University of Michigan Museum of Zoology, http://animaldiversity.org/accounts/Acipenser_transmontanus/.

Fiske, Julia A., et al. "A Comparison of Methods for Determining Ploidy in White Sturgeon (Acipenser Transmontanus)." *Aquaculture*, vol. 507, 2019, pp. 435–42, https://doi.org/10.1016/j.aquaculture.2019.03.009

Jelena Rajkov, Zhaojun Shao, Patrick Berrebi "Evolution of Polyploidy and Functional Diploidization in Sturgeons: Microsatellite Analysis in 10 Sturgeon Species." *Journal of Heredity*, vol. 105, no. 4, 2014, pp. 521–531, https://doi-org.offcampus.lib.washington.edu/10.1093/jhered/esu027.

Krieger, Jeannette, et al. "Phylogenetic Relationships of the North American Sturgeons (Order Acipenseriformes) Based on Mitochondrial DNA Sequences." *Molecular Phylogenetics and Evolution*, vol. 16, no. 1, 2000, pp. 64–72, https://doi.org/10.1006/mpev.1999.0743

Mitofish: Mitochondrial Genome Database of Fish. Atmosphere and Ocean Research Institute, the University of Tokyo, 2021, http://mitofish.aori.u-tokyo.ac.jp/. Accessed 9 Jun. 2020.

Nelson R.J. & McAdam D.S.O. "Historical population structure of White Sturgeon in the Upper Columbia River detected with a combined analysis of capture, telemetry and genetics." *Journal of Applied Ichthyology*, vol. 28, no. 2, 2012, 161–167, https://doi-org.offcampus.lib.washington.edu/10.1111/j.1439-0426.2011.01930.x

Rastorgeuv S., Mugue N., Volkov A. & Barmintsev V. "Complete mitochondrial DNA sequence analysis of Ponto-Caspian sturgeon species." *Journal of Applied Ichthyology*, vol. 24, no. 1, 46-49, https://doi-org.offcampus.lib.washington.edu/10.1111/j.1439-0426.2008.01089.x.

Shen, Yanjun, et al. “Phylogenetic Perspective on the Relationships and Evolutionary History of the Acipenseriformes.” *Genomics*, 2020, pp. 1–7, https://doi.org/10.1016/j.ygeno.2020.02.017

Teresa L. Gilbert, James R. Brown, Patrick J. O’Hara, Norman E. Buroker, Andrew T. Beckenbach, Michael J. Smith. “Sequence of tRNA Thr and tRNA Pro from white sturgeon (Acipenser transmontanus) mitochondria.” *Nucleic Acids Research*, vol. 16 no. 24, 1988, pp. 11825, https://doi.org/10.1093/nar/16.24.11825

Welcome to Python.org. Python Software Foundation, 2021, https://www.python.org/. Accessed 9 Jun. 2020.