Mitochondrial Phylogenomics versus Nuclear-Based Phylogenies of Poison Frogs: NGS Mitogenomes COX1-Derived Barcodes, and Their Comparison with Nuclear Gene Phylogenies

Authors: Ruby Lin[1], Ravindra Managar[1] Mentor: Juan C Santos[3]

[1]Stuyvesant High School; [2]Santos Lab, St. John's University

Abstract: Using mitogenomes derived from RNAseq data of 10 different poison frog species and 4 outgroups, we estimated a phylogeny based on whole COX1 sequences derived from these RNA sequencing data. For this purpose, total RNA of four species were sent to for Next-Generation Sequencing and a bioinformatics pipeline help us to derive the mitochondrial genomes, annotate these sequences, and isolate the COX1 gene. Our host lab (Santos lab) provided us with whole COX1 gene sequences of 10 additional species and we also obtained available nuclear sequences from the NCBI (GenBank) database. We compare the COX1 phylogeny with the one derived from the combined nuclear genes. For this purpose, we analyzed and contrasted both topologies, their nodal supports, and tip (species) placements. We found that the whole-gene COX1 phylogeny performed better than the one derived from nuclear-genes and this COX1 phylogeny agrees with published results on these frogs. Our results highlight how the whole COX1 could be derived from RNAseq data and why COX1 does a good job in reconstructing phylogenies of our focal amphibians.

Introduction

The mitochondrial genome consists of DNA within the mitochondrion, containing genes crucial for cellular respiration processes. Its size is generally around 16,500 base pairs, and it is subject to a high mutation rate. The mitochondrial genome contains the COX1 gene used for barcoding in the molecular identification and systematics in a wide variety of species. This DNA is unique because it usually follows a maternal inheritance pattern and can, therefore, be used to determine maternal lineages. Mitochondrial DNA is also in the form of a plasmid consisting of circular DNA that runs antiparallel to each other [2]. Its size is generally around 16,500 base pairs, and it is subject to a high mutation rate [1]. Several reasons behind such increase include its proximity where cellular respiration occurs, its "naked" nature that exposes mitochondrial DNA to mutagenic agents, and its poor or lack of reparation mechanisms. Within the mitochondrial DNA, the COX1 gene encodes the cytochrome c oxidase subunit I of the complex IV of the mitochondrial respiratory chain and this protein plays a fundamental role in energy production of aerobic cells. Sequences of this gene have been used as barcodes of life because it is easy to amplify using near-universal COX1 primers from the ubiquitous mtDNA that can be extracted from almost any tissue [3]. After extraction and amplification, the DNA needs to be sequenced using traditional Sanger sequencing. Since 2010's, Next-Generation Sequencing or NGS has revolutionized DNA or RNA sequencing where resulting information can range from whole genomes to transcriptomes. Furthermore, RNA-Seq data tend to produce large quantities of byproduct organelle sequences (e.g., mitochondrial and chloroplast DNA) that can also be used to recover entire mitochondrial genomes [4]. Therefore, NGS data can easily replace standard COX1 sequencing and some of the nuclear genes revealed by NGS can also be used to construct phylogenies based on nuclear genes. By using NGS data to construct the whole mitochondrial genome, we can extract the whole COX1-barcode gene and compare it to a nuclear gene-based phylogenies. Using this approach, we can reveal discrepancies on the evolutionary position of species as we catalog biodiversity of our planet using molecular phylogenetics.

Materials & Methods

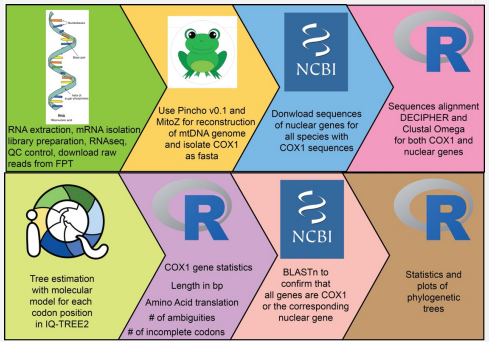

Table 1. Length of COX1 per species

Taxa	COX1 length (bp)
Allobates insperatus	1545
Allobates kingsburyi	1545
Allobates talamancae	1545
Ameerega bilinguis	1545
Colostethus ruthveni	1542
Dendrobates vanzolinii	1554
Eleutherodactylus johnstonei	1548
Eleutherodactylus planirostris	1557
Hyloxalus azureiventris	1557
Ikakogi tayrona	1545
Leucosthetus fugax	1542
Paruwrobates erythromos	1503
Pristimantis viejas	1551
Rheobates palmatus	1542

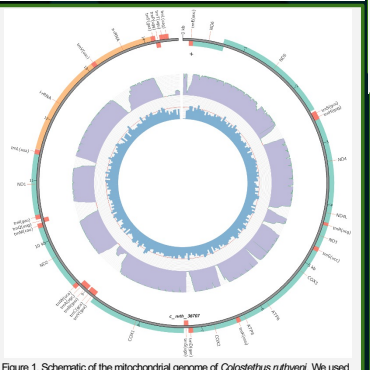

Figure 1. Schematic of the mitochondrial genome of *Colostethus ruthveni*. We used COX1 ("barcode gene") for further analyzes. Notice the absence of cytochrome b gene and D-loop, this artifact could be improved with additional sequencing.

Table2. NCBI accession numbers of nuclear gene per species.

Taxa	RAG1	RHO	TYR
Allobates insperatus	--	--	HQ290899
Allobates kingsburyi	--	--	HQ290901
Allobates talamancae	DQ503363	DQ503250	HQ290912
Ameerega bilinguis	--	DQ503216	HQ290933
Colostethus ruthveni	MF614351	MF614392	MF624155
Dendrobates vanzolinii	DQ503318	DQ503204	--
Eleutherodactylus johnstonei	--	--	--
Eleutherodactylus planirostris	--	--	EF493470
Hyloxalus azureiventris	--	--	HQ290915
Ikakogi tayrona	EU663456	--	--
Leucosthetus fugax	--	--	HQ290898
Paruwrobates erythromos	--	--	MW042075
Pristimantis viejas	--	--	JN991593
Rheobates palmatus	--	MF614424	HQ290905

Results and Discussion

We extracted total RNA from muscle tissue of 4 species of frogs: *Pristimantis viejas*, *Colostethus ruthveni*, *Paruwrobates erythromos*, and *Ikakogi tayrona*. After quantification of extractions, samples were sent to sequence to Azenta, Inc. (NJ) for standard RNA sequencing (RNAseq). Briefly, total RNA integrity was evaluated using Bioanalyzer and quantified prior to Figure 1. Schematic of the mitochondrial genome of *Colostethus ruthveni*. We used COX1 ("barcode gene") for further analyzes. Notice the absence of cytochrome b gene and D-loop, this artifact could be improved with additional sequencing or modification of MitoZ [12] parameters. mRNA isolation. All four samples had a RIN >7.5, which indicated that sample had a good integrity, mRNA was isolated using poly-A selection, and samples could proceed to RNAseq library preparation. Samples were pair-end sequenced (2x150bp) using the Illumina platform. Total raw reads were *P. viejas* (29,500,760 reads), *C. ruthveni* (29,844,486), *P. erythromos* (26,823,939), and *I. tayrona* (27,623,401). The files were sent via ftp and they were processed using a de novo transcriptome assembly approach using Pincho v0.1 [13] with default parameters. For the mitochondria annotation and visualization, we used MitoZ [12] under the following parameters: An insertion size of 150 bp, minimum abundance equal to 3 and Chordate as clade of reference. Other parameters were for default using 12 threads. An example of partial mt-genome is provided in Figure 1.

We found that the RNAseq provide enough sequencing data to the complete COX1 gene as evidenced with the alignment provided in our results. We found that the whole-gene COX1 phylogeny performed better than the one derived from nuclear-genes for the frog species studied. We compared with published reports and the COX1 phylogeny was similar to those with more genes (e.g., mitochondrial: 12S, 16S rRNA, ND1, ND2, CYTB; nuclear: RAG1, RHO, TYR, SIA). RNA-Seq experiments tend to produce large quantities of byproduct mitochondrial sequences that can also be used to recover to reconstruct entire mitochondrial genome. We think that NGS approaches might potentially replace standard COX1 sequencing. Our results highlight how the whole COX1 could be derived from RNAseq data and why COX1 does a good job in reconstructing phylogenies of our focal amphibians.

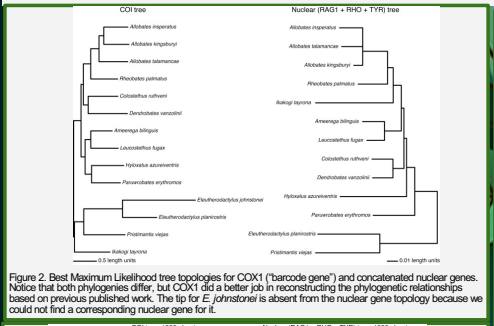

Figure 2. Best Maximum Likelihood tree topologies for COX1 ("barcode gene") and concatenated nuclear genes. Notice that both phylogenies differ, but COX1 did a better job in reconstructing the phylogenetic relationships based on previous published work. The tip for *E. johnstonei* is absent from the nuclear gene topology because we could not find a corresponding nuclear gene for it.

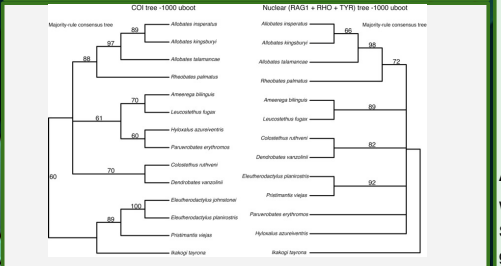

Figure 3. Bootstrap tree topologies for COX1 ("barcode gene") and concatenated nuclear genes. Notice that COX1 has a better nodal support (i.e., >70) that the nuclear genes. The tip for *E. johnstonei* is absent from the nuclear gene topology because we could not find a corresponding nuclear gene for it.

References

1. Nature Publishing Group. (n.d.). Nature news. Retrieved November 29, 2021, from https://www.nature.com/scitable/topicpage/mitos-and-mitochondrial-diseases-903/.
2. *Mitochondrial DNA*. Genome.gov. (n.d.). Retrieved November 29, 2021, from https://www.genome.gov/genetics-glossary/Mitochondrial-DNA.
3. Rodrigues, M. S., Morelli, K. A., & Jansen, A. M. (2017, October 16). *Cytochrome c oxidase subunit 1 gene as a DNA barcode for discriminating trypanosoma cruzi dtus and closely related species*. Parasites & Vectors. Retrieved November 29, 2021, from https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-017-2457-1.
4. Forni, G., Puccio, G., Bourguignon, T., Evans, T., Mantovani, B., Rota-Stabelli, O., & Luchetti, A. (2019, October 15). *Complete mitochondrial genomes from transcriptomes: Assessing pros and cons of data mining for assembling new mitogenomes*. Nature News. Retrieved November 29, 2021, from https://www.nature.com/articles/s41598-019-51319-7.
5. Paradis E. & Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526-528.
6. D.F. Robinson and L.R. Foulds (1981) Comparison of phylogenetic trees, Mathematical Biosciences, 53(1), 131–147
7. Penny, D. and Hendy, M. D. (1985) The use of tree comparison metrics. Systematic Zoology, 34, 75–82.
8. Kuhner, M. K. and Felsenstein, J. (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates, Molecular Biology and Evolution, 11(3), 459–468
9. Smith, M.R. (2020a). Information theoretic Generalized Robinson-Foulds metrics for comparing phylogenetic trees. Bioinformatics 36: 5007–5013. doi:10.1093/bioinformatics/btaa614
10. D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, and L.S. Vinh (2018) UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol., 35:518–522. https://doi.org/10.1093/molbev/msx281
11. Grant, T., M. Rada, M. Anganoy-Criollo, A. Batista, P. H. Dias, A. M. Jeckel, D. J. Machado, and J. V. Rueda-Almonacid. 2017. Phylogenetic systematics of dart-poison frogs and their relatives revisited (Anura: Dendrobatoidea). South American Journal of Herpetology 12:S1-S90.
12. Meng, G., Y. Li, C. Yang, and S. Liu. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res 47:e63.
13. Ortiz, R., P. Gera, , C. Rivera, J. Santos, C, and 2021. Pincho: A modular approach to high quality de novo transcriptomics. . Genes 12:953; https://doi.org/910.3390/genes12070953.
14. Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Z. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:6.
15. Minh, B., Q. H. Schmidt, A. O. Chernomor, D. Schrempf, M. Woodhams, D, A. von Haeseler, and R. Lanfear. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530-1534.

Acknowledgements

We thank the Urban Barcoding Research Project (UBRP) and the Santos Lab. Allison Mayle and Arden Feil provided mentorships, guidance, and support. We thank The Pinkerton Foundation and Science Sandbox for their financial support.