

# The effects of ecological traits on the rate of molecular evolution in ray-finned fishes: a multivariable approach

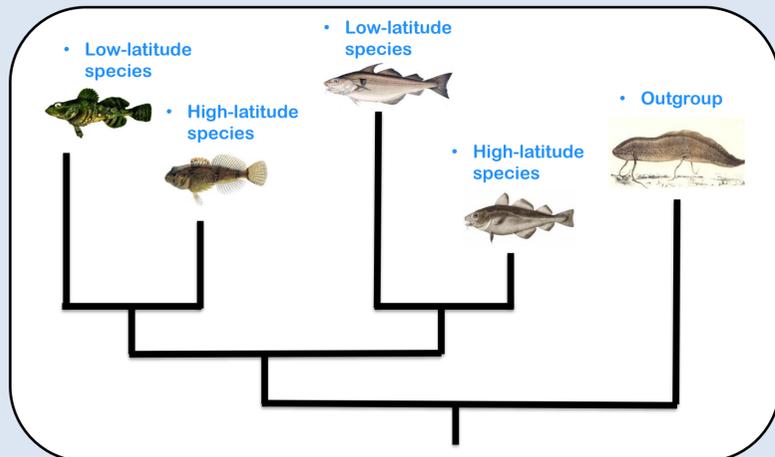
Jacqueline A. May, Zeny Feng, Matthew G. Orton, Sarah J. Adamowicz

## Abstract

**Background:** A myriad of environmental and biological traits have been investigated for their roles in influencing the rate of molecular evolution. However, most studies have focused on a single ecological trait, while controlling for additional factors in an informal way, generally by excluding taxa. The public availability of DNA barcode sequence data, in conjunction with online trait databases, provides the opportunity to perform a more comprehensive study that considers the effects of multiple traits at a broader phylogenetic scale and in a multivariable context. **Results:** This study utilized a dataset comprised of cytochrome c oxidase subunit I (COI) barcode sequences from over 6000 fish species to investigate the effects of 32 biological and environmental traits on molecular evolutionary rates. Latitude and environmental temperature were included, which have been previously implicated as correlates of mutation rates in fish, as were traits related to population structure, life history, and habitat. Additional traits previously associated with metabolic rate differences in fish, such as morphology, locomotion, and feeding habits, were also investigated. A bioinformatics pipeline in the R programming language was constructed to manipulate and assemble both DNA barcode data retrieved from the BOLD API and trait data obtained from FishBase. Results from our regression analyses, accounting for phylogeny, revealed molecular evolutionary rates to correlate most significantly with biological traits such as age at maturity and longevity. **Significance:** These results provide evidence for the importance of life history traits as molecular rate correlates in fishes relative to environmental traits, while also showcasing the efficiency of using bioinformatics tools to assemble and analyze biological information obtained from different online databases. The bioinformatics pipeline may be easily adapted to investigate additional organismal groups and molecular rate correlates, thereby providing a solid and amendable foundation for future work in this field.

## Introduction

- Ecological traits have been investigated for their role in influencing the **rate of molecular evolution**. However, many studies are limited to a single trait, and few<sup>1,2</sup> have considered multiple traits.
- Molecular rate may be measured by comparing the **substitution rate** of organisms that differ in a certain trait (Fig. 1).
- The objective of this study is to construct a **bioinformatics pipeline** that investigates the effects of **multiple** ecological traits on the rate of molecular evolution in fish (including both previously investigated and novel traits).

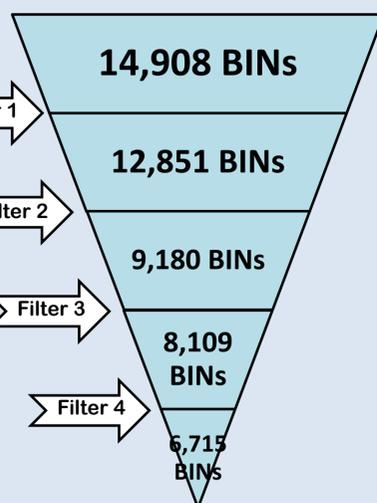


**Figure 1:** In this hypothetical example, latitude (as a proxy for temperature) is expected to correlate with molecular evolutionary rates in ectothermic organisms such as fishes.

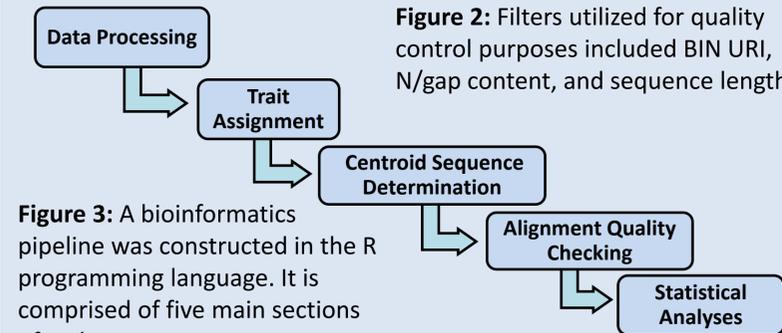
Mottled Sculpin by Joe Tomelleri, Shorthorn Sculpin from arctic.uoguelph.ca; Arctic Cod from ian.umces.edu, Haddock by Gervais et. Boulart (1877); Lungfish by G.H. Ford.

## Materials & Methods

- Fish COI<sup>3</sup> sequences downloaded from **Barcode of Life Data System (BOLD)**<sup>4</sup> database.
- Barcode Index Numbers (BINs)**<sup>5</sup> used as proxy for species.
- Data for 32 traits obtained from **FishBase**<sup>6</sup>.
- Rates** of molecular evolution quantified as **branch lengths**.
- Single-variable and multivariable regression analyses performed using traits as **predictors** and branch length as **response** variable.



**Figure 2:** Filters utilized for quality control purposes included BIN URI, N/gap content, and sequence length.

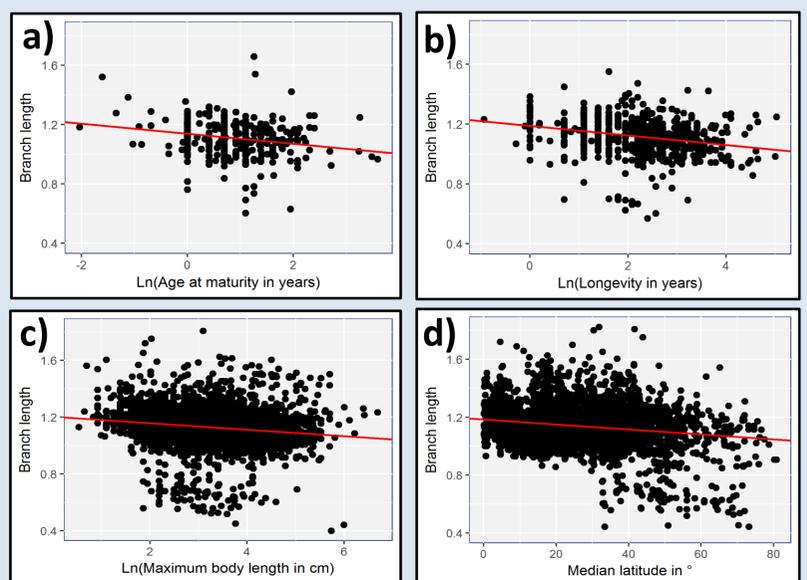


**Figure 3:** A bioinformatics pipeline was constructed in the R programming language. It is comprised of five main sections of code.

## References

- Bromham, L., Hua, X., Lanfear, R., & Cowman, P.F. (2015). Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. *Am. Nat.* **185**: 507-524.
- Fujisawa, T., Vogler, A.P., & Barraclough, T.G. (2015). Ecology has contrasting effects on genetic variation within species versus rates of molecular evolution across species in water beetles. *Proc. R. Soc. B.* **282**: 1-9.
- Hebert, P.D., Cywinska, A., Ball, S.L., & deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. B.* **270**: 313-321.
- Ratnasingham, S., & Hebert, P.D.N. (2007). BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Mol. Ecol. Notes.* **7**: 355-364.
- Ratnasingham, S., & Hebert, P.D.N. (2013). A DNA-based registry for all animal species: the Barcode Index Number (BIN) System. *PLoS ONE.* **8**: e66213.

## Results



**Figure 4:** Select single-variable regression results showing the effects of a) age at maturity ( $n = 331$ ,  $p$ -value  $< 0.0001$ ,  $R^2 = 0.25$ ), b) longevity ( $n = 663$ ,  $p$ -value  $= 0.0005$ ,  $R^2 = 0.19$ ), c) maximum body length ( $n = 3339$ ,  $p$ -value  $< 0.001$ ,  $R^2 = 0.12$ ) and d) latitude ( $n = 4763$ ,  $p$ -value  $= 0.0004$ ,  $R^2 = 0.09$ ) on branch lengths estimated using all three codon positions and controlling for phylogeny and number of nodes.

**Table 1:** Best-fit multivariable models after backwards selection process ( $n = 175$ ), accounting for phylogeny. Bold p-values indicate significance at 0.05 level.

Trait	Whole codon		Third codon	
	Estimated Coefficient (B)	P-Value	Estimated Coefficient (B)	P-Value
Number of Nodes	0.0092	<b>&lt; 0.0001</b>	0.015	<b>&lt; 0.0001</b>
Age at Maturity	-0.0034	<b>0.0012</b>	-0.0095	<b>0.0015</b>
Maximum Water Depth	0.000010	0.24	0.000017	0.50
Median Latitude	-0.00025	0.43	-0.00045	0.62
Maximum Body Length	-0.000049	0.51	-0.000024	0.91

## Conclusions

- The strongest predictor of molecular rate was **age at maturity**, which provides evidence for the **generation-time effect**<sup>7</sup> in fishes.
- A non-significant **latitudinal** effect in the multivariable analyses suggests that **temperature-driven** molecular rates do not appear to be a **universal attribute** across ectothermic groups.
- These results provide evidence for the importance of **life history** rate correlates relative to **environmental** factors in fish COI.
- Relaxed clock methods** that permit lineage rate variation would likely be appropriate at a broad phylogenetic scale when analyzing COI fish sequences.
- The **bioinformatics pipeline** designed for this study may be adapted to investigate trait correlates in novel taxonomic groups.

## References continued

- Frøese, R., & Pauly, D. (Eds.) (2016). FishBase. World Wide Web electronic publication. [www.fishbase.org](http://www.fishbase.org).
- Ohta, T. (1993). An examination of the generation-time effect on molecular evolution. *PNAS.* **90**: 10676-10680.

## Funding Sources/Acknowledgements

This research was funded by NSERC Discovery Grants to Dr. Sarah J. Adamowicz and Dr. Zeny Feng. Special thanks to Dr. Robert Hanner for providing advice on how to obtain and process barcode and trait data.

**Github link to pipeline:** <https://github.com/jmay29/phylo>.