



## RECOMMENDATION

# An open-source pipeline to reconstruct phylogenies with paleoproteomic data

Leslea J Hlusko<sup>1</sup> 

<sup>1</sup> Centro Nacional de Investigación sobre la Evolución Humana (CENIEH) – Burgos, Spain

**Cite as:** Hlusko LJ (2023). An open-source pipeline to reconstruct phylogenies with paleoproteomic data. *Peer Community in Paleontology*, 100220. DOI: 10.24072/pci.paleo.100220

**Published:** 19th September 2023

**Based on reviews by:**  
Katerina Douka and two anonymous reviewers

**Correspondence:**  
leslea.hlusko@cenieh.es

© 2023 Author(s)



This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.

## A recommendation of

Patramanis I, Ramos-Madrigal J, Cappellini E, and Racimo F (2023). PaleoProPhyler: a reproducible pipeline for phylogenetic inference using ancient proteins. *bioRxiv* 519721, ver. 3 peer-reviewed by PCI Paleo. DOI: 10.1101/2022.12.12.519721

One of the most recent technological advances in paleontology enables the characterization of ancient proteins, a new discipline known as palaeoproteomics (Ostrom et al., 2000, Warinner et al., 2022). Palaeoproteomics has superficial similarities with ancient DNA, as both work with ancient molecules, however the former focuses on peptides and the latter on nucleotides. While the study of ancient DNA is more established (e.g., Shapiro et al., 2019), palaeoproteomics is experiencing a rapid diversification of application, from deep time paleontology (e.g., Schroeter et al., 2022) to taxonomic identification of bone fragments (e.g., Douka et al., 2019), and determining genetic sex of ancient individuals (e.g., Lugli et al., 2022). However, as Patramanis et al. (2023) note in this manuscript, tools for analyzing protein sequence data are still in the informal stage, making the application of this methodology a challenge for many new-comers to the discipline, especially those with little bioinformatics expertise.

In the spirit of democratizing the field of palaeoproteomics, Patramanis et al. (2023) developed an open-source pipeline, PaleoProPhyler released under a CC-BY license (available on GitHub, see Appendix). Here, Patramanis et al. (2023) introduce their workflow designed to facilitate the phylogenetic analysis of ancient proteins. This pipeline is built on the methods from earlier studies probing the phylogenetic relationships of an extinct genus of rhinoceros *Stephanorhinus* (Cappellini et al., 2019), the large extinct ape *Gigantopithecus* (Welker et al., 2019), and *Homo antecessor* (Welker et al., 2020). PaleoProPhyler has three interacting modules that initialize, construct, and analyze an input dataset. The authors provide a demonstration of application, presenting a molecular hominid phyloproteomic tree.

In order to run some of the analyses within the pipeline, the authors also generated the Hominid Palaeoproteomic Reference Dataset which includes 10,058 protein sequences per individual translated from publicly available whole genomes of extant hominids (orangutans, gorillas, chimpanzees, and humans) as well as some ancient genomes of Neanderthals and Denisovans. This valuable research resource is also publicly available, on Zenodo (Patramanis et al., 2022).

Three reviewers reported positively about the development of this program, noting its importance in advancing the application of palaeoproteomics more broadly in paleontology.

## References

- Cappellini E, Welker F, Pandolfi L, Ramos-Madrigal J, Samodova D, R  ther PL, Fotakis AK, Lyon D, Moreno-Mayar JV, Bukhsianidze M, Rakownikow Jersie-Christensen R, Mackie M, Ginolhac A, Ferring R, Tappen M, Palkopoulou E, Dickinson MR, Stafford TW, Chan YL, G  therstr  m A, Nathan SKSS, Heintzman PD, Kapp JD, Kirillova I, Moodley Y, Agusti J, Kahlke RD, Kiladze G, Mart  nez-Navarro B, Liu S, Sandoval Velasco M, Sinding MHS, Kelstrup CD, Allentoft ME, Orlando L, Penkman K, Shapiro B, Rook L, Dal  n L, Gilbert MTP, Olsen JV, Lordkipanidze D, and Willerslev E (2019). Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature* 574, 103–107. DOI: 10.1038/s41586-019-1555-y.
- Douka K, Brown S, Higham T, P  abo S, Derevianko A, and Shunkov M (2019). FINDER project: collagen fingerprinting (ZoomS) for the identification of new human fossils. *Antiquity* 93, e1. doi: 10.15184/aqy.2019.3.
- Lugli F, Nava A, Sorrentino R, Vazzana A, Bortolini E, Oxilia G, Silvestrini S, Nannini N, Bondioli L, Fewlass H, Talamo S, Bard E, Mancini L, M  ller W, Romandini M, and Benazzi S (2022). Tracing the mobility of a Late Epigravettian (~ 13 ka) male infant from Grotte di Pradis (Northeastern Italian Prealps) at high-temporal resolution. *Scientific Reports* 12, 8104. doi: 10.1038/s41598-022-12193-6.
- Ostrom PH, Schall M, Gandhi H, Shen TL, Hauschka PV, Strahler JR, and Gage DA (2000). New strategies for characterizing ancient proteins using matrix-assisted laser desorption ionization mass spectrometry. *Geochimica et Cosmochimica Acta* 64, 1043–1050. doi: 10.1016/S0016-7037(99)00381-6.
- Patramanis I, Ramos-Madrigal J, Cappellini E, and Racimo F (2022). Hominid Palaeoproteomic Reference Dataset. doi: 10.5281/ZENODO.7333226.
- Patramanis I, Ramos-Madrigal J, Cappellini E, and Racimo F (2023). PaleoProPhyler: a reproducible pipeline for phylogenetic inference using ancient proteins. *bioRxiv* 519721, ver. 3 peer-reviewed by PCI Paleo. doi: 10.1101/2022.12.12.519721.
- Schroeter ER, Cleland TP, and Schweitzer MH (2022). Deep Time Paleoproteomics: Looking Forward. *Journal of Proteome Research* 21, 9–19. doi: 10.1021/acs.jproteome.1c00755.
- Shapiro B, Barlow A, Heintzman PD, Hofreiter M, Pajmans JLA, and Soares AER, eds. (2019). *Ancient DNA: Methods and Protocols*. 2nd ed. Vol. 1963. Methods in Molecular Biology. New York, NY: Humana. isbn: 978-1-4939-9175-4. doi: 10.1007/978-1-4939-9176-1.
- Warinner C, Korzow Richter K, and Collins MJ (2022). Paleoproteomics. *Chemical Reviews* 122, 13401–13446. doi: 10.1021/acs.chemrev.1c00703.
- Welker F, Ramos-Madrigal J, Gutenbrunner P, Mackie M, Tiwary S, Rakownikow Jersie-Christensen R, Chiva C, Dickinson MR, Kuhlwilm M, De Manuel M, Gelabert P, Martin  n-Torres M, Margvelashvili A, Arsuaga JL, Carbonell E, Marques-Bonet T, Penkman K, Sabid   E, Cox J, Olsen JV, Lordkipanidze D, Racimo F, Lalueza-Fox C, Berm  dez De Castro JM, Willerslev E, and Cappellini E (2020). The dental proteome of *Homo antecessor*. *Nature* 580, 235–238. doi: 10.1038/s41586-020-2153-8.
- Welker F, Ramos-Madrigal J, Kuhlwilm M, Liao W, Gutenbrunner P, De Manuel M, Samodova D, Mackie M, Allentoft ME, Bacon AM, Collins MJ, Cox J, Lalueza-Fox C, Olsen JV, Demeter F, Wang W, Marques-Bonet T, and Cappellini E (2019). Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. *Nature* 576, 262–265. doi: 10.1038/s41586-019-1728-8.

## **Appendix**

Reviews by Katerina Douka and two anonymous reviewers, DOI: 10.24072/pci.paleo.100220.

PaleoProPhyler on GitHub: [https://github.com/johnpatramanis/Proteomic\\_Pipeline](https://github.com/johnpatramanis/Proteomic_Pipeline)