

# tombRaider – improved species and haplotype recovery from metabarcoding data through artefact and pseudogene exclusion.

Dr Gert-jan Jeunen<sup>1</sup>, Dr Kristen Fernandes<sup>1</sup>, Dr Eddy Dowle<sup>1</sup>, Gracie Kroos<sup>1</sup>, Dr Quentin Mauvisseau<sup>2</sup>, Dr Michal Torma<sup>2</sup>, Dr Allison Miller<sup>1</sup>, Prof Miles Lamare<sup>1</sup>, Prof. Neil Gemmell<sup>1</sup>

<sup>1</sup>University Of Otago, <sup>2</sup>University of Oslo

## Biography:

Research Fellow at the University of Otago

## Abstract:

Environmental DNA metabarcoding has revolutionized ecological surveys of natural systems. By amplifying and sequencing small gene fragments from environmental samples containing complex DNA mixtures, scientists are now capable of exploring biodiversity patterns across the tree of life in a time-efficient and cost-effective manner. However, the accuracy of species and haplotype identification can be compromised by sequence artefacts and pseudogenes. Despite various strategies developed over the years, effective removal of artefacts remains challenging and inconsistent data reporting standards hinder reproducibility in eDNA metabarcoding experiments. To address these issues, we introduce tombRaider, an open-source command line software program (<https://github.com/gjeunen/tombRaider>) and R package ([https://github.com/gjeunen/tombRaider\\_R](https://github.com/gjeunen/tombRaider_R)) to remove artefacts and pseudogenes from metabarcoding data post clustering and denoising. tombRaider features a modular algorithm capable of evaluating multiple criteria, including sequence similarity, co-occurrence patterns, taxonomic assignment, and the presence of stop codons. We validated tombRaider using various published data sets, including mock invertebrate communities, air eDNA from a zoo, and salmon haplotypes from aquatic eDNA. Our results demonstrate that tombRaider effectively removed a higher proportion of artefacts while retaining authentic sequences, thus enhancing the accuracy and reliability of eDNA-derived diversity metrics. This user-friendly software program not only improves data quality in eDNA metabarcoding studies, but also contributes to standardised reporting practices, an aspect currently lacking in this emerging research field.

## ABOUT eDNA

The Southern environmental DNA Society (SeDNA) is a newly established Australian and New Zealand society of environmental DNA researchers and end users. We aim to promote best practices and help the adoption of methods across sectors.

Our mission is promoting science and industry collaboration across Australia and New Zealand to advance best practice eDNA methods and adoption in government, private and community sectors.

Visit our website to find more about the society and what we do here. Membership registration is open on our website.

<https://sednasociety.com/>

## CONFERENCE MANAGERS

Please contact the team at Conference Design with any questions regarding the conference.

[mail@conferencedesign.com.au](mailto:mail@conferencedesign.com.au)

+61 3 6231 2999

[www.conferencedesign.com.au](http://www.conferencedesign.com.au)

## LINKS

[Code of Conduct](#)

[Privacy and Data Policy](#)

## ACKNOWLEDGEMENT OF COUNTRY

In the spirit of reconciliation we acknowledge the Traditional Custodians of country throughout Australia & New Zealand and their connections to land, sea and community. We pay our respect to their Elders past and present and extend that respect to all peoples today.