

**eDNA Metabarcoding: Gaps of Publicly Available Reference Databases of Freshwater Vertebrates in Sri Lanka**

**Ranaweera J.B.S.<sup>1\*</sup>, Fernando T.S.P.<sup>2</sup>, Weerakoon D.K.<sup>1</sup>**

<sup>1</sup>*Department of Zoology and Environmental Science, University of Colombo, Colombo 03, Sri Lanka*

<sup>2</sup>*Department of Zoology, Open University, Nawala, Sri Lanka*

\**judyranaweera@gmail.com*

**Abstract**

Environmental DNA (eDNA) has become a rapidly emerging technology during the last decade. Although eDNA applications are used worldwide, none of the eDNA metabarcoding studies have been conducted solely in Sri Lanka yet. Detection rate of species through DNA metabarcoding technique varies with the availability of gene sequences in reference databases. Therefore, the main objective of this study was to assess the data gaps of local species in two selected reference databases (the NCBI database and the BOLD database). Local freshwater vertebrates were considered for this analysis by dividing them into three major categories as freshwater fish (97 species), freshwater amphibians (40 species) and freshwater reptiles (7 species). The availability of COI gene sequence and CytB gene sequence of each species were recorded. In the NCBI database, reference gene sequences were available only for 17.53% (17 species) of freshwater fish, 7.50% (3 species) of freshwater amphibians and 14.29% (1 species) of freshwater reptiles. In the BOLD database, reference gene sequences were available for 52.58% (51 species) of freshwater fish, 17.50% (7 species) of freshwater amphibians and 42.86% (3 species) of freshwater reptiles. Endemic species were analyzed separately. This revealed the availability of reference gene sequences for 4.92% (3 species) of endemic freshwater fish in the NCBI database. However, none of the reference gene sequence was found for endemic freshwater amphibians and reptiles. In the BOLD database, reference gene sequences were available for 36.07% (22 species) of endemic freshwater fish and 33.33% (1 species) of endemic freshwater reptiles but none for endemic freshwater amphibians. According to the results, large gaps of publicly available reference COI gene sequences were observed in both NCBI and BOLD databases. Not only the availability of reference gene sequences, but also the submitted countries of those sequences were also analyzed as a part of this study. Asia has contributed the most according to the country-wise analysis. Among them, many are in the South Asian region. Therefore, this study highlights the need of expanding reference databases for greater local representation as well as the importance of forging collaborations with the developed countries in the same region in order to increase international cooperation and funding to carry on further studies on neglected taxa and under-represented taxa in Sri Lanka.

**Keywords:** eDNA (environmental DNA), Metabarcoding, Reference databases, Reference gene sequences