

Appendix J. DNA Barcoding and Metabarcoding Explained

DNA barcodes are short, species specific genetic sequences used to identify taxa, similar to how a fingerprint can identify individual humans or the way a supermarket scanner can distinguish products using the black stripes of the Universal Product Code. For metazoans (organisms from the animal kingdom), the leading genomic region for DNA barcoding is the cytochrome oxidase I (COI) gene of the mitochondrial genome. Prior to DNA barcoding, biological specimens were identified using morphological features such as color, size, and shape of body parts and required taxonomic training, the use of morphological “keys”, and undamaged specimens. DNA barcoding solves these problems by requiring just a tiny amount of tissue and no taxonomic expertise.

Barcoded organisms tend to be individually photographed, preserved, and when possible, visually assigned to the lowest known taxon. Once sequenced, they are termed Operational Taxonomic Units (OTU) and are typically matched at a 97% sequence similarity to existing sequences with known species names that reside within publicly available DNA-barcode libraries such as the Barcode of Life Data Systems (BOLD; <http://www.boldsystems.org/>). OTUs are pragmatic proxies for “species” and those that match directly with known sequences within the DNA libraries are assigned a species name. In the absence of a direct match, OTUs can undergo a phylogenetic computational approach to assign them to higher taxonomic levels such as to a genus or family level. As more and more organisms are vouchered, identified, barcoded, and their sequences submitted to DNA-barcode libraries, previously unidentified sequences can regularly be blasted (cross-checked) against the barcode libraries to increase taxon resolution assuming the sequences are stored and properly archived.

DNA metabarcoding is a cutting-edge molecular technique that resembles DNA barcoding in that it targets a specified gene and obtains short genetic sequences for identification. However, this technique sequences hundreds to thousands of organisms at one time rather than focusing on a single unique individual. Thus, individual metabarcoding samples are viewed as sampled communities.

Metabarcoding samples tend to be bulked, homogenized, or filtered (water) and may or may not have obvious signs of biological material. As a result, barcode libraries are critical to metabarcoding techniques in the identification of organisms. Given the sheer diversity of coral reef cryptobiota, there are thousands of organisms that do not have an associated DNA barcode identified in a library and are, in fact, probably new to science. Thus, to date, species-level identification is not possible for all sequences obtained from metabarcoding. Sequences not identified can be clustered into OTUs based on DNA sequence similarity as discussed above. As DNA libraries grow, sequences can be re-examined to not only obtain species identifications but to investigate community composition metrics. Innovative bioinformatic (computational) processing mechanisms focused on the phylogenetic relationships of sequences are currently being developed and will provide and enhance a deeper resolution of the millions of sequencing data obtained from combined ARMS units to understand diversity and composition across spatial scales. As these methods develop and improve, the ARMS deployed around Timor-Leste will be integrated into a global analysis of cryptobiota diversity and composition, the first of its kind.