BUILDING A REFERENCE DNA BARCODE LIBRARY FOR MARINE AMPHIPODS OF PORTUGAL

FERREIRA, M. S.[1], ARTEAGA, J. L.[2], GOMES, P. A.[1], CUNHA M. R. [3], COSTA, M. H.[2], COSTA, F.O.[1]

1-Centre of Molecular and Environmental Biology (CBMA), Department of Biology/ University of Minho, Campus de Gualtar 4710-057 Braga, Portugal; 2- IMAR - Instituto do Mar, Universidade Nova de Lisboa, Campus de Caparica, 2829-516 Caparica, Portugal. 3- Centro de Estudos do Ambiente e do Mar, Departamento de Biologia, Universidade de Aveiro, 3810-193 Aveiro, Portugal

Usefulness and validity of data generated in amphipod monitoring programmes depends on ease of access to rigorous and expedite taxonomic identifications. DNA barcodes of the mitochondrial gene cytochrome oxidase 1 (CO1) have been shown to clearly discriminate species in most animal groups, including amphipod crustaceans. We initiated the creation of a reference DNA barcode library (RDBL) for marine amphipods of Portugal in order to provide a tool to assist monitoring programmes and improve taxonomic rigor in biodiversity inventories. Here we describe the analytical chain leading to the creation of a RDBL, stressing the unique aspects of this approach and the associated data quality control procedures.

All specimens collected for RDBL were preserved in 96% ethanol and upon completion of the study will be archived in the National Museum of Natural History. Specimens were morphologically identified to species level, and quality images acquired. Metadata from the specimen's collection event (e.g. GPS coordinates) were lodged in a public database (BOLD). CO1-barcodes were gathered by means of DNA extraction, PCR amplification and sequencing, following published protocols. Complementary strands of each sample were edited and aligned manually, and uploaded to BOLD. To rank as a DNA barcode standard, sequences must have a minimum sequence length, obtained from bidirectional sequences, and sequence trace files must be

available for checking. To inspect the taxonomic congruence of the barcode data, we analysed sequence divergence, clustering patterns in a NJ tree, and performed homology searches against GenBank and BOLD databases. This approach enabled us to find important ambiguities between our data and data from GenBank, most of them lacking specimen images or trace files that could enable quality control and effective data comparison. The key element lies in establishing an annotation system for single sequence data in public databases enabling taxonomic re-examinations.