ICES/PICES 6ZPS 2016/W7

Evaluating DNA-based Identifications of Next-Generation Sequencing and Classical Sanger Sequencing Using a Comprehensive 18S rDNA Reference Library of the North Sea Metazoan Fauna

Inga Mohrbeck[1], Michael J. Raupach[1], Pedro Martínez Arbizu[1], Thomas Knebelsberger[1], Silke Laakmann[1]

[1] Senckenberg am Meer, German Center for Marine Biodiversity Research (DZMB), Südstrand 44, 26382 Wilhelmshaven, Germany

Next-generation sequencing has been proposed as an efficient tool to document biodiversity on a regional or greater scale. In this framework, we tested the applicability of high-throughput molecular identification using next generation pyrosequencing in marine metazoans. A comprehensive reference library of the 18S rDNA: V1-V2 region, based on Sanger sequencing, was constructed for 118 North Sea species from a wide range of taxonomic groups to directly compare DNA-based identifications of 454 sequencing with traditional morphological identification methods. The chosen gene region was evaluated for its effectiveness in species delineation by clustering the reference library at different sequence similarity thresholds. The optimal threshold was found at 99% similarity, with 85% identification success. Pyrosequencing of the V1-V2 region was then used to analyze replicates of two samples containing DNA pooled from the previously Sanger-sequenced species. The 454-derived sequences were compared to our reference library to test if the generated reads would represent all of the species included in the pooled samples. Pyrosequencing was able to identify between 67% and 78% of the species only. Also, a large number of sequences for three species that were not included in the pooled samples were amplified by pyrosequencing. The study demonstrated that, based on the V1-V2 region, 454 sequencing does not provide accurate species differentiation and reliable taxonomic classification. The analysis of artificially prepared samples indicated that species detection in pyrosequencing datasets is complicated by potential PCR-based biases and that the V1-V2 marker is poorly resolved for some taxa.

Key words: Sequence databases, Reference library, Next- generation sequencing, 454 pyrosequencing, Biodiversity assessment

Contact author: Inga Mohrbeck, Senckenberg am Meer, German Center for Marine Biodiversity Research (DZMB), email: imohrbeck@senckenberg.de