

Investigating the marine invertebrate fauna of the West African continental shelf with DNA barcodes

Endre Willassen, Jon A. Kongsrud, Katrine Kongshavn, Manuel A.E. Malaquias, Tom Alvestad

University Museum of Bergen, Department of Natural History, P.O. Box 7800, 5020 Bergen, Norway

Abstract

We have sorted and identified invertebrates from about 700 sampling stations distributed from Morocco to Angola. In joint efforts with African, European, and American partners we have prepared about 2600 samples of predominantly Polychaeta, crustacean Malacostraca, Gastropoda, and Echinodermata for barcoding with the BOLD facilities. Despite relatively low sequencing success in some of the animal groups, DNA-barcodes are giving us a wealth of new insights into the biodiversity of the African Atlantic. For some of the more well studied animal groups with DNA barcodes already present in open access databases positive barcode identifications has given us new information about species distributions, species associations, life history stages and genetic population structure. DNA barcodes are also revealing taxonomic discordance amongst identifiers in different laboratories and indicates the need for taxonomic revision of many database records.

Background

The University Museum of Bergen, Norway, has been cooperating with the EAF-Nansen-Project (<http://www.fao.org/in-action/eaf-nansen/en>) since 2005 in exploring benthic invertebrate diversity on the continental shelf of the Canary Current (CCLME) and Gulf of Guinea Large Marine Ecosystems (GCLME). We have sorted and identified samples from about 700 sampling stations distributed from Morocco to Angola. In joint efforts with African, European, and American partners we have prepared about 2600 samples of predominantly Polychaeta, crustacean Malacostraca, Gastropoda, and Echinodermata for barcoding with the BOLD facilities (Figure 1).

Sequencing results

Relatively high proportions, on average 60 %, of the samples prepared for sequencing did not produce barcode compliant results. Low success rates as in for example some of the polychaete groups seem to be due to “universal primer” mismatch. In many cases we missed the genome of target species due to competing DNA from other organisms in the sample environment. These contaminations could be bacteria, parasites or even sometimes other animals more closely related to the target species.

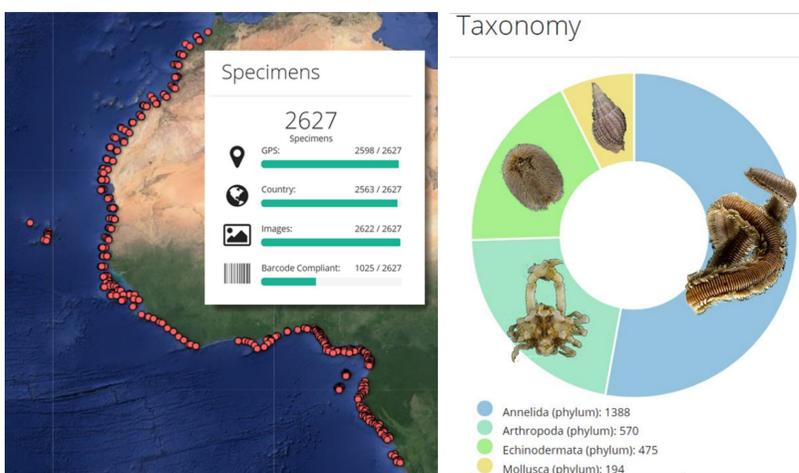


Figure 1: Map of sampling sites with barcoded specimens and screen dumps from BOLD showing MIWA project efforts and results.

Positive barcode identifications

A growing DNA library of identified COX1 sequences from other barcoding initiatives has provided alternative means to traditional morphology based species identification. By using information in BOLD and Genbank we have observed several cases of surprising species occurrences that considerably expand previously known distributions. Some of these cases are particularly striking because they exemplify how identifications can still be confidently achieved from animals and animal remains even when diagnostic morphological characteristics are missing or vague (Figure 2).



Figure 2: *Astrospartus mediterraneus* (left) and *Cellana cf. torreuma* (right). Examples of unexpected species occurrences discovered by matching sequences in the BOLD database; the former species primarily known from the Mediterranean Sea, the latter from the Indian Ocean (see Willassen et al. 2016).

References

- <https://miwa.w.uib.no/>
Malaquias MAE, Ohnheiser LT, Oskars TR, Willassen E. 2016 Diversity and systematics of philineid snails (Gastropoda: Cephalaspidea) in West Africa with remarks on the biogeography of the region. *Zoological Journal of the Linnean Society*. doi: 10.1111/zoj.12478
Willassen E, Williams AB, Oskars TR. 2016. New observations of the enigmatic West African *Cellana* limpet (Mollusca: Gastropoda: Nacellidae). *Marine Biodiversity Records*. doi: 10.1186/s41200-016-0059-9.

Taxonomic impacts and challenges

DNA barcodes provide independent data to validate taxonomic decisions based on traditional morphological methods. Consequently, barcodes may reveal many cases of taxonomic discordance that may be rooted either in misidentifications or in inappropriate or underdeveloped taxonomies for the group in question. The BOLD databases have become an important forum for biodiversity discoveries, a trigger for taxonomic enquiries and a driver for new systematic studies.

• Identical sequences – different species names

BOLD data present many cases of taxonomic discordance that may reflect different taxonomic cultures in different countries, different uses of identification literature, different notions about species, or plain misidentifications by individual workers.

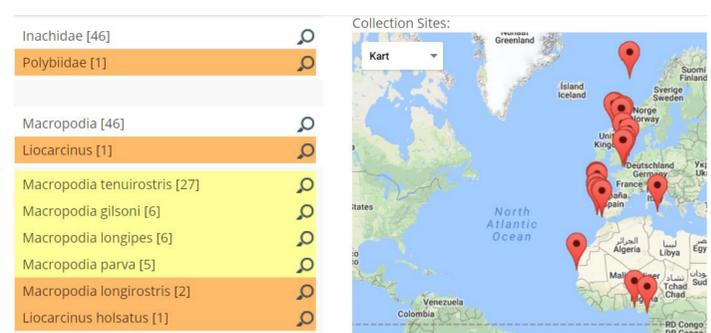


Figure 3: Screen dump from BOLD showing six conflicting species identifications for similar sequences by different identifiers working in different fauna regions.

• Same species name – different sequences

For many of the less well studied groups there is an overwhelming genetic diversity among gene tree clades that were initially presumed to be one species. A large number of “morpho-species” have been assigned to several different BINs in BOLD, and are considered candidates for new species recognitions. Interestingly, some of such highly divergent clades with sometimes more than 20 % p-distance have sympatric geographic occurrence and may seem to be sometimes structured by depth, temperature and water masses more than geographical distances.

Table 1: *Scalibregma inflatum* (Annelida: Scalibregmatidae) – a morphologically defined species with highly divergent COX1. Many similar examples of genetic divergence have been recorded.

BIN	# spms	Max. intra. variation	Distance to nearest BIN
AAC6624	10	1.24 %	>4.8% (ABU2370)
AAF2370	4	1.00 %	>2.8% (ABU2370)
ABU6064	13	0.46 %	>2.8% (AAF2370)
ACI7945	3	0.00 %	>29% (ACA1331)
ACA1331	2	0.31 %	>26% (AAC6624)
ACI8225	10	1.31 %	>23% (AAC6624)



• No species name – different sequences

Some of the macro-benthos of W. Africa are relatively well known from historical inventories and more recent studies. However, for several of the species rich and less studied groups much of the material has no recognizable species descriptions in relevant literature. We are observing previously undiscovered species and for some groups we are expecting even a ten-fold increase in species numbers. We also try to produce DNA barcodes for as yet undescribed species and to have new species analyzed and described within an integrative framework by taxonomic specialists of the relevant group (Malaquias et al. 2017).

Funding

Specimen sampling was funded by the EAF-Nansen program. Sample sorting, identification, curation etc. was supported by the University of Bergen with support from JRS Biodiversity Foundation. The Canadian Center of DNA Barcoding and BIO Guelph indirectly subsidized the barcoding activities through Boldsystems.org.

Acknowledgements

Thanks are due to the international group of colleagues who study biodiversity and have invested their time in parts of this project.