

Barcode of Life: Advancing Species Identification and Discovery

D. Chandramohan

Consultant, National Institute of Oceanography, Kochi & Member, International Scientific Steering Committee, Census of Marine Life(CoML), Washington, D.C
Email: drd.chandramohan@gmail.com

As we know the taxonomic expertise is getting reduced dangerously and the species diagnosis still remains as a major challenge for any biological research. In one of the papers published by Hebert (Hebert et al., 2003) he highlights the daunting task of identifying millions of species when compared to the physicists who deal with a cosmos assembled from 12 fundamental particles. Since few taxonomists can identify more than 0.01% of the estimated 10-15 million species, a community of 15,000 taxonomists will be required continuously to identify life forms on this planet if we want to rely on the morphological diagnosis as we practice now. The problem becomes more severe when we deal with identification of larval stages of animals in sea water. Larval dispersal and colonization are key stages in the life history of planktonic spawning marine invertebrates. The direct study of these processes requires the accurate identification of often microscopic larvae, which in many cases lack diagnostic morphological characters particularly when dealing with the earliest developmental stages. Many researchers have successfully used few molecular techniques such as RAPD (Randomly Amplified Polymorphic DNA), RFLP (Restriction Fragment Length Polymorphism) species specific DNA probes etc. Dr. Hebert further points out four major limitations of these classical methods of identification:

1. Both phenotypic plasticity and genetic variability in the characters employed can lead to incorrect identification.
2. DNA barcoding approach overlooks morphologically cryptic taxa, which are seen very commonly in many groups.
3. Since morphological keys are often effective only for a particular life stage or gender, many individuals in their early stages of development can not be properly identified.
4. Although modern interactive identification keys represent a major advance, such keys often require a high level of expertise and lack of such expertise may even lead to wrong identifications.

These limitations in morphology-based identification systems and the dwindling pool of taxonomists highlight the need for alternate methods for species identification which should be quick, cost effective and efficient. DNA barcoding emerges as a most favoured alternate method by the researchers.

DNA barcoding is a technique that uses a short gene sequence from a standardized region of the genome as a diagnostic “biomarker” for the species. Different species have different DNA barcodes making it possible to use barcodes to (i) identify specimens, (ii) discover possible new species and (iii) make taxonomy more effective for science and society. Can we generate such unique codes for these millions of species? The Universal Product Codes, used to identify retail products, employ 10 alternate numerals at 11 positions to generate

100 billion unique identifiers. Genomic barcodes have only four alternate nucleotides (A, G, C, T) at each position, but the string of sites available for inspection is huge. The survey of just 15 of these nucleotide positions creates the possibility of 4^{15} (1 billion) codes, 100 times the number that would be required to discriminate life on earth if each taxon was uniquely branded.

Studies conducted so far indicate that a DNA barcode consisting of a 648-bp region near the 5' end of the mitochondrial Cytochrome-C oxidase subunit I (COI) gene provides a highly effective system for the identification and discovery of animal life. It is now clear that the mitochondrial genome of animals is a better target for analysis than the nuclear genome because of its lack of introns, its limited exposure to recombination and its haploid mode of inheritance (Saccone et.al., 1999). Robust primers also enable the routine recovery of specific segments of the mitochondrial genome (Folmer et.al, 1994; Simmons and Weller, 2001). Past phylogenetic work has often focused on mitochondrial genes encoding ribosomal (12S, 16S) DNA, but their use in broad taxonomic analyses is constrained by the prevalence of insertions and deletions (indels) that greatly complicate sequence alignments (Doyle and Gaut, 2000). The 13 protein-coding genes of the mitochondria appear to be the best targets as the occurrence of 'indels' are rare. Even though there is no proper unique reason for selecting COI, it does have two important reasons: (i) universal primer for this gene is very robust and works well with most of the animal phyla and (ii) possess greater range of phylogenetic signal than any other mitochondrial gene.

The accumulated evidence now shows that these short DNA sequences can be a distinguishing feature from insects to birds. As a Linnaean binomial is an abbreviated label for the morphology of a species, the short sequence is an abbreviated label for the genome of the species. Thus, the Barcode of Life (BOL) provides additional master key to the knowledge about a species. Hebert and his co-workers (Hebert et.al., 2003) created a COI 'profile' for 200 closely allied species and subsequently used it to assign 150 newly analyzed individuals to species proving that this method proved 100% successful in identification. Further, the application of this approach has already permitted the delineation of previously unrecognized species of both vertebrates and invertebrates. However, while animal barcode records are accumulating at an accelerating rate, work on other groups of life has barely been initiated. Researchers are now trying to extend DNA barcoding to plants, fungi and protists. Considerable advancements have been made at least with two major animal groups – birds and fishes. Currently four major initiatives are in operation:

1. Lepidoptera : This initiative will assemble barcodes for 25,000 species of Lepidoptera with a focus on the faunas of Australia, Canada, Costa Rica and United States of America.
2. Fishes :The FISH-BOL programme will gather barcodes for at least 15,000 fish species with an emphasis on marine species(<http://www.fishbol.org/>)
3. Birds: The All-Birds Barcode initiative (ABBI) plans to assemble DNA Barcodes for all 10,000 bird species within 5 years.
4. Marine Zooplankton: The ongoing Census of Marine Life (CoML) project, Census of Marine Zooplankton (CmarZ)-<http://www.cmarz.org/>- will barcode identified specimens of any animal holoplanktonic species.

As on today, there are 1,65,084 barcodes are available primarily in the Databases of GenBank (25,723) and Guelph Facility at Canada (1,34,990) which account for 19,163 species.

The Consortium for the Barcode of Life (CBOL) (<http://www.barcodinglife.org>) is an international collaboration of natural history museums, herbaria, biological repositories and biodiversity inventory sites together with academic and commercial experts in genomics, taxonomy, electronics and computer science. The mission of the CBOL is to develop DNA barcoding as a global standard in taxonomy, rapidly accelerate compiling of DNA barcodes of known and newly discovered plant and animal species, establish a public library of sequences linked to named specimens and promote development of portable devices for DNA barcoding. CBOL has more than 130 member organizations from 40 countries. The CBOL has also developed the Barcode of Life Database (BOLD) which is a repository for barcode records and also has analytical tools which can serve as an online workbench for the DNA barcode community (<http://www.barcodinglife.org>).

The CBOL gives ten reasons for why this technology should be promoted.

1. Works with fragments : Barcoding can identify a species from bits and pieces.
2. Works for all stages of life : Barcoding can identify a species in its many forms from eggs and seed, through larvae and seedlings, to adults and flowers.
3. Unmasks look-alikes : Barcoding can distinguish among species that look alike, uncovering dangerous organisms masquerading as harmless ones and enabling a more accurate view of biodiversity.
4. Reduce ambiguity : A library of digital barcodes will provide an unambiguous reference that will facilitate identifying species invading and retreating across the globe and through centuries.
5. Makes expertise go further : Scientists can equip themselves with barcoding to speed identification of known organisms and facilitate rapid recognition of new species.
6. Democratizes access : A standardized library of barcodes will empower many more people to call by name the species around them.
7. Opens the way for an electronic handheld field guide, the "Life Barcoder"
8. Sprouts new leaves on the Tree of Life (TOL) : Barcoding newly discovered species will help show where they belong among known species, sprouting new leaves on the tree of life.
9. Demonstrates value of collections: Compiling the library of barcodes begins with the multi millions of specimens in museums, herbaria, zoos and gardens. The spotlight that barcoding shines on these institutions and their collections will strengthen their ongoing efforts to preserve Earth's Biodiversity.
10. Speeds writing the encyclopedia of life : Compiling a library of barcodes linked to vouchered specimens and their binomial names will enhance public access to biological knowledge, helping to create an on-line encyclopedia of life on Earth, with a web page for every species of plant and animal.

Researchers have already applied this technology to the fouling organisms especially the Barnacles. The DNA barcodes using the COI gene sequence for the following species have already been submitted to the GenBank database: *Balanus amphitrite*, *B.glandula*, *Chthamalus stellatus*, *C.improvisus*, *C.eburneus*, *C.crenatus*, *C.malayensis*, *C.proteus*, *C.dalli*,

C.panamensis, *C.fragilis*, *C.cortezianus*, *C.mexicanus*, *C.challengeri* etc. Obviously this technology would become more useful to identify the early developmental stages of the organisms in natural environment and subject them to test the efficacy of antifouling compounds under laboratory conditions.

As the cost of sequencing is dropping rapidly due to technical advances and the existence of high throughput systems, processing of large number of samples in short time has become a reality. Once reference database is established, this technology can be applied even by non-specialist.

References

1. Doyle, J.J. & Gaut, B.S. 2000 Evolution of genes and taxa: a primer. *Plant Mol. Biol.* **42**, 1-6.
2. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R., 1994 DNA primers for amplification of mitochondrial cytochrome oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **3**, 294-299.
3. Herbert, D.N.P., Alina Cywinska, Shelly L. Ball and Jeremy R. de Waard. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B.* **270**, 313-321.
4. Saccone, C., Decarla, G., Gissi, C., Pesole, G. and Reyes, A. 1999. Evolutionary genomics in the Metazoa: the mitochondrial DNA as a model system. *Gene* **238**, 195-210.
5. Simmons, R.B. and Weller, S.J. 2001. Utility and evolution of cytochrome *b* in insects. *Mol. Phylogenet. Evol.* **20**, 196-210.