

DNA barcodes

"We are convinced that the sole prospect for a sustainable identification capability lies in the construction of systems that employ DNA sequences as taxon 'barcodes'." (Hebert et al. 2003)

How it is proposed to work:

A short sequence, ~650bp, potentially contains enough information to resolve 10-100million species. Now “micro barcodes” of 100bp are being proposed for more degraded materials.

A good DNA barcode sequence is conserved enough to be amplified with “universal” primers while divergent enough to resolve closely related species. COI is presumed to have these properties.

1. Gather this short sequence from all samples.
2. Build "profile" trees. Generally NJ is used.
3. Match taxonomic names to terms.
4. For unknowns their identity can be read from the resultant topology, typically, but not always by grouping with a cluster that is 98% or more similar.

Purported good properties and possible applications:

1. Rapid and accurate identifications.
2. Rapid enumeration of species.
3. ~1.7 million species described and at least 10 million others are yet to be named. DNA-based identification will allow for flagging specimen that probably represent undescribed taxa, e.g. “cryptic species” (cynically defined as specimens with divergent COI haplotypes that barcoders want someone else to test and describe as species)
4. When it becomes cheap, it will be available to local non-experts via portable barcode machines. Proponents claim this is the answer to the problem of the lack of “Bioliteracy”.

Problems:

-Resolving recently diverged species and hybrids may be impossible for COI and there is no way to know when the answer is wrong except in well known and well sampled groups. The right length of sequence data needed in any given group to recover correct results is not known and probably will vary across taxa. Must be able to distinguish between interspecific and intraspecific variation.

-No single gene is conserved across all life. So it will take a few at least.

-Reference sequences must be from “taxonomically confirmed” specimens or one must accept unique COI haplotype clusters as the “important” units. This is Hebert’s gene-species and Blaxter’s MOTUs.

-*Identification does not equal Science*: Identification is placing individuals into circumscribed groups or recognizing that they fit none that exist. The science, taxonomy and systematics, determining those groups would have to be nearly finished for this to even be partially useful beyond tentatively identifying samples.

- A taxon gives the characters the characters do NOT give the taxon

- After systematic and taxonomic study (integrative taxonomy) the best tools to identify important biological units can be used. It might be DNA. Might even be COI.

-How is divergence in COI related to species boundaries and what is its relationship to units for which we have interesting biological questions? Fundamentally, barcoding ignores difficult issues of species concepts and boundaries, dynamics of biological classifications and taxonomic hypotheses. In the BOLD system... “*Reference barcodes are a validated subset of the full database containing only those species represented by three or more individuals showing less than 2% sequence divergence.” Is this a barcode species?

Mitochondrial heteroplasmy

Identical sequences in different species

Introgression and Hybrid speciation

Incomplete lineage sorting

NUMTs- Nuclear Pseudogenes

-*It does not answer the Biodiversity crisis*: species diversity; ecological cohesion, evolutionary relationships. Its applicability to conservation biology and biodiversity surveys for rapid and accurate identifications is wholly dependent on current knowledge and expert help.

-A shift in focus and money would be devastating to biology as a whole, as it does not reduce the need for taxonomists only delays the hard work and reduces the potency of the "encyclopedia of life".



