OVERVIEW ON TAXONOMY WITH DNA BARCODING

Hüseyin Özdikmen* and Hakkı Taştan*

* Gazi University, Science Faculty, Department of Biology, 06500 Ankara, TURKEY. E-mails: ozdikmen@gazi.edu.tr; hakkitastan@gazi.edu.tr

[Özdikmen, H. & Taştan, H. 2020. Overview on taxonomy with DNA barcoding. Munis Entomology & Zoology, 15 (2): 627-630**]**

ABSTRACT: DNA barcoding is a new an important key for species identification and has greatly carried out the pace of species discovery. In this aspect, more effective technique, a short DNA sequence from a standard region of mitochondrial " called " barcode" is used for taxonomy field. Barcode sequence has been employed as a possible DNA marker for species and a number of studies in a variety of taxa have accordingly been performed to examine its efficacy. In some studies, barcoding provided a means of highlighting potential cryptic, synonymous or extinct species as well as matching adults with immature specimens. However, higher taxa have not been resolved as accurately as spercies . Despite everything, DNA barcoding appears to offer a means of identifying species and may become a standart tool. Depart from here, recent developments in high throughput parallel sequencing technology, combined with DNA barcoding , provide an advance that can bring together the feilds of phylogenetic and phlogeographic analysis to dissolve any problem between species. Such an approach could be standardized for analyses at any geographical scale for a range of taxonomic groups to quantify the formation and composition of species assemblages. This review article explores some of these issues, focusing on taxonomy and DNA barcoding.

KEY WORDS: DNA barcoding, taxonomy, overview

As used in many areas, digitization has a positive effect on biodiversity work. Museums get the order of As, Gs, Cs and Ts strings from DNA sequencing machines and turn these physical analog samples into digital objects. Libraries and commercial publishers can convert physical books and articles into images, as well as convert them to letter strings using optical character recognition (OCR) systems. When these developments are analyzed, there are striking similarities between the formation of DNA sequence databases in the twentieth century and the rise of natural history museums in previous centuries (Strasser, 2008, 2011).

Converting both digital taxonomy and genomic sequence to digital form is very important for taxonomic studies. Some difficulties may be encountered while doing these conversion jobs. Some of these difficulties may be: algorithms developed for binary order alignment have applications in extracting articles from OCR text (Page, 2011). Looking at these challenges from a different perspective, how to handle the extraordinary growth of data for Sequence databases, for taxonomy, it is about how to make an immigrant in a large number of non-digital representations (Lathe et al., 2008; Arinõ, 2010).

Barcode studies seem to be very important in overcoming these difficulties. By identifying common barcode sequences for the species and transferring them to digital media, a database is created for researchers who will work with similar types (Lathe et al., 2008; Arinõ, 2010; Page, 2011).

It is suggested that this approach is guiding in the taxonomic studies related to that type by determining the barcode sequences of the species and putting them into digital environment (Prendini, 2005; Arinõ, 2010).

ISSN 1306-3022

DNA BARCODE APPROACH IN TAXONOMY

One of the major issues related the inclusion of molecular information into taxonomic aspects of biology that has yet to be debated in detail in the commentaries on this subject is related the best way to read the barcodes. There are two separate tasks to which DNA barcodes are being applied right now. The first is the use of DNA data to differenciate between species (equivalent to species identification or species diagnosis) and the second is the use of DNA data to determine new species (equivalent to species delimitation, species description). These two activities differ in the types and amount of data required. in this chapter; it is highlighted some of the issues that may limit the utility of current DNA barcoding efforts (especially those used for species discovery) and proposal a framework for the development of a barcoder that addresses these issues.

(a) With a barcode-based database: distances or characters?

With the realization of barcode creation, one of the important issues is how to read this barcode. In the most recently published articles, it has been stated that distance measurements are made between the species by using DNA barcoding in the identification of the species (Hebert et al., 2003; Prendini, 2005). For distance measurements between species, two basic logics are used; The first is a simple BLAST (Altschul et al., 1990) approach, in which a raw similarity value is used to identify the neighbor closest to the query string. The second basic logic is to make tree construction by making use of the blasted distances (Hebert et al., 2003). Accordingly, most of the data suggest that these two basic, and also character-based approaches are more suitable for DNA barcoding for both theoretical and practical reasons.

(b) With the database based on barcode research:

The question arises that barcoding studies using character-based information that are not related to the tree or accordingly are a viable alternative to distance-based approaches applied today. With this question, it turns out that the approach to analysis of which characters belong to which species (non-tree population aggregation analysis or a tree-based approach) is more suitable for barcoding (Nixon & Wheeler, 1990). There may be some disadvantages to using tree forming approaches in species identification. The first is related to the use of distances (as described above) to build trees. However, problems with tree construction are not limited to trees obtained by using distance data. On the contrary, the second disadvantage is the use of single gene trees as evidence of phylogenetic relationships. Many studies on this subject have demonstrated that the combined analysis of multiple data segments, theoretically (Kluge, 1989) and empirically (Gatesy et al., 1999; Rokas et al., 2003), represents the evolutionary process better than single gene trees. The combined analysis approach is to investigate the character contribution of data segments to the combined tree and separate analysis of individual gene segments is thought to elicit character support for the combined tree, which is not exactly evident (Baker & DeSalle, 1997; Baker et al., 1998; Gatesy et al., 2002, 2003). From this perspective, it can be argued that a verified total evidence tree can be used as a guide tree to describe the phylogenetic state of the sequence of an unknown species, assuming that the query string is one of the gene regions used to construct the total evidence tree. There are bioinformatics tools for solving problematic points in the tree, which will help determine a query sequence based on the presence of shared characters identified in the species (Sarkar et al., 2002).

(c) With barcode research database: how many types are enough?

Another controversial aspect of the DNA barcoding initiative relates to the number of individuals of each putative species to include in the analysis. Classical taxonomy attempts to come to the conclusion by scanning a large number of individuals to identify characters shared uniquely among all members of a genre and to distinguish variation in a genre from variation between species. In such studies, sometimes one or more people may not represent the species as a whole, especially among the widely distributed taxa (Nixon & Wheeler, 1990; Walsh, 2000; Goldstein & DeSalle, 2003). The requirement for a sufficient number of individuals applies to both distance and character-based methods, and it may not be likely to have a universal sample size suitable for all species. There is also no universal geographical distance to provide a reasonable proxy to determine the appropriate sampling strategy. In the light of all this information, as in the selection of gene regions, taking samples representing a sufficient number of species to capture the in-species representative variation, in the pilot studies to be carried out on them, other information about these species, as well as their life history, distribution ability and mating models. may need to be used as plan information (Walsh, 2000).

RESULTS

DNA barcoding stands out as an important system for identifying types based on this information, allowing users to link samples to taxonomic information databases and access some information about species that do not yet have data. It is very important to reveal some important information about species at the molecular level with a barcoding approach. However, it should be emphasized that this situation is not exactly a taxonomic explanation or phylogenetic structure. However, barcode studies can speed up the work of taxonomists and other people interested in type identification. To date, it is known that the results obtained from the studies with barcode have an important place in the system related to the identification of the species and its location in taxonomy.

DISCUSSION

In scientific studies, both taxonomy and barcoding actively digitize the living world. The identification of new animal taxa is basically proceeding at a fixed rate. With the digitization of these taxa, taxonomic literature has an increasing level of growth potential. The use of DNA characters in the identification of species is completely in line with the current taxonomic research process. When the studies on barcode are examined, it is determined that the results obtained from barcode studies are consistent and compatible with the results obtained from current taxonomic applications. Taxonomy and barcode studies have also been found to support each other in the identification and classification of species by acting as a bridge between morphological and molecular approaches. In this respect, it is also thought that it is an effective method for differentiation of the species belonging to the Dorcadionini tribe (Coleoptera: Cerambycidae) and the identification of some new species that are considered to belong to this tribe with the barcode study approach. We conclude that non-tree approaches are more suitable for the construction of a barcode reader for the following reasons. First, tree-based approaches can produce phylogeny based on a poorly chosen molecule for phylogenetics. Although trees generally make sense, support for hypotheses from such trees is almost always low and may limit the validity of phylogenetic

ISSN 1306-3022

hypotheses from such trees (Kluge, 1989; Rokas et al., 2003; Gatesy et al., 2003). Based on this, the results of the relationship between species on a phylogenetic tree created by a single molecular marker will contradict existing approaches to modern systematics. Second, current taxonomic approaches perform tree-independent diagnosis to create taxonomic systems. The use of DNA characters in the diagnostic context for species identification is known to be fully compatible with the current taxonomic research process. Third, the recommendations based on the literature are also consistent with current taxonomic practices. It has been determined that it is possible to differentiate the types and sometimes to introduce new types with barcode applications which are frequently used recently.

In a review article with barcode studies, it has also been seen that dna barcode applications have been used quite frequently recently, and it is an important molecular approach in determining species separation and their place in classification.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol., 215: 403-410 (doi:10.1006/jmbi.1990.9999.).

Arinõ, A. H. 2010. Approaches to estimating the universe of natural history collections data. Biodivers. Inf., 7 (doi:10.17161/bi.v7i2.3991).

Baker, R. H. & DeSalle, R. 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. Syst. Biol., 46: 654-673.

Baker, R. H., Yu, X. & DeSalle, R. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol., 9: 427-436 (doi:10.1006/mpev.1998.0519.).

Gatesy, J., Amato, G., Norell, M., DeSalle, R. & Hayashi, C. 2003 Combined support for wholesale taxic atavism in gavialine crocodylians. Syst. Biol., 52: 403-422.

Gatesy, J., Matthee, C., DeSalle, R. & Hayashi, C. 2002. Resolution of a supertree/supermatrix paradox. Syst. Biol., 51: 652-664 (doi:10.1080/10635150290102311.).

Gatesy, J., O'Grady, P. & Baker, R. H. 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artio-dactyl taxa. Cladistics, 15: 271-313 (doi:10.1111/j.1096-0031.1999.tb00268.x.).

Goldstein, P. Z. & DeSalle, R. 2003. Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. Mol. Ecol., 12: 1993-1998 (doi:10. 1046/j.1365-294X.2003.01860.x.).

Hebert, P. D. N., Ratnasingham, S. & deWaard, J. R. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. B, 270 (Suppl.): 96-99.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool., 38: 7-25.

Lathe, W., Williams, J., Mangan, M. & Karolchik, D. 2008. Genomic data resources: challenges and promises. Nat. Educ., 1, 2.

Nixon, K. C. & Wheeler, Q. D. 1990. An amplification of the phylogenetic species concept. Cladistics, 6: 211-223.

Page, R. D. M. 2011. Extracting scientific articles from a large digital archive: BioStor and the Biodiversity Heritage Library. BMC Bioinformatics, 12: 187 (doi:10.1186/1471-2105-12-187).

Prendini, L. 2005. Comment on 'Identifying spiders through DNA barcoding'. Can. J. Zool., 83: 498-504 (doi:10.1139/z05-025.).

Rokas, A., Williams, B. L., King, N. & Carroll, S. B. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature, 425: 798-804 (doi:10. 1038/nature02053.).

Sarkar, I. N., Thornton, J. W., Planet, P. J., Figurski, D. H., Schierwater, B. & DeSalle, R. 2002. An automated phylogenetic key for classifying homeoboxes. Mol. Phylogenet. Evol., 24: 388-399 (doi:10.1016/S1055-7903(02) 00259-2.).

Strasser, B. J. 2008. GenBank – natural history in the 21st century? Science, 322: 537-538 (doi:10.1126/science.1163399).

Strasser, B. J. 2011. The experimenter's museum: GenBank, natural history, and the moral economies of biomedicine. Isis, 60-96 (doi:10.1086/ 658657).

Walsh, P. 2000. Sample size for the diagnosis of conservation units. Conserv. Biol., 14: 1533-1535 (doi:10.1046/j.1523-1739.2000.98149.x.).