

# Current perspectives of DNA barcoding in Bosnia and Herzegovina through BOLD database

Trenutne perspektive DNK barkodiranja u Bosni i Hercegovini putem BOLD baze podataka

Lejla Ušanović<sup>1,\*</sup>, Jasna Hanjalić<sup>1</sup>, Semir Dorić<sup>1</sup>, Lejla Lasić<sup>1</sup>, Jasmina Šubara<sup>2</sup>, Adnan Čučuković<sup>2</sup>, Belma Kalamujić Stroil<sup>1</sup>

<sup>1</sup> University of Sarajevo-Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina

<sup>2</sup> Society for Genetic Conservation of Bosnia and Herzegovina's Endemic and Autochthonous Resources - GENOFOND, Sarajevo, Bosnia and Herzegovina

## ABSTRACT

DNA barcoding is a method designed to provide rapid and precise species identifications by using one or more of short gene sequences called barcodes. In most plant and fungi studies, the standard barcodes of choice are three plastid (*rbcl*, *matK* and *trnH-psbA*) and one nuclear (*ITS*) gene regions. The relatively high, but comparatively conserved rate of sequence evolution of mtDNA has made *COI* the marker of choice in animals. BOLD is a freely available cloud-based data storage and analysis platform developed with the aim to advance biodiversity science through DNA barcoding species identification. To date, over 6 million barcodes have been deposited in BOLD with 196,000 animal species, 68,000 plant species and 22,000 species of fungi and other organism entries. In this database, there are currently 447 entries for organisms from Bosnia and Herzegovina, which makes 0.0067% of the total number of BOLD entries. According to BOLD statistics, only 1.11% of all organism entries from B&H were submitted by B&H institutions.

Despite the fact that Bosnia and Herzegovina has valuable natural resources with a high percentage of endemic and autochthonous species, BOLD statistics elucidated the lack of coordinated and systematic DNA barcoding research so far. It is necessary to establish continuous progress of molecular-genetic characterization of these resources in the future. It is up to B&H institutions to decide if they want to continue the practice of submitting the data sporadically or if they will animate the research community to actively participate in this global project.

**Key words:** biodiversity databases, BOLD Systems, DNA barcoding

## INTRODUCTION - Uvod

In the era of the “taxonomic crisis” when only a small number of taxonomists can reliably identify species by applying morphological methods, the establishment of a

fast, reliable and cost-effective tool for species identification becomes a global need. This operational problem was the driving force for the development of a new method - DNA barcoding, designed to enable rapid and precise identification of species using one or more

\* Corresponding author: Lejla Ušanović, University of Sarajevo-Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina; e-mail address: lejla.usanovic@ingeb.unsa.ba

short sequence genes from the entire genome, the so-called barcodes (Hebert et al. 2003a). DNA barcodes (400-800 bp) should be easily generated, informative and applicable for all species on the planet (Kress & Erickson, 2008). In most of the previous studies, three plastid gene regions (*rbcl*, *matK* and *trnH-PsbA*) and one nuclear (*ITS*) were used as standard barcodes for plants and fungi, respectively. The DNA sequence for cytochrome c oxidase subunit I (*COI*) fits DNA barcodes criteria and has been used as the marker of choice in animals (Kress & Erickson, 2012). The DNA barcode concept was described in 2003 by P.D.N. Hebert and associates from the Guelph University of Ontario, who used primers to amplify 648 bp of *COI* gene, described by Folmer et al. (1994). The idea of barcoding began with the creation of a “*COI* profile” for the representatives of different species from 100 families of eight orders of insect classes. The authors proposed this mitochondrial gene as the basis of the molecular bioidentification system in animals that provides reliable, economical and accessible solutions to current problems in identifying species (Hebert et al. 2003a; Hebert et al. 2003b).

DNA barcoding involves two basic steps: (1) building a reference sequence library of barcodes of already known species, and (2) associating barcodes with sequences of unknown patterns with barcodes from the database. DNA barcoding is useful in cases when species identification by a classic morphological method is hindered, e.g. samples in different developmental stages (Kalamujić Stroil et al. 2018a), damaged samples, specimens from the gut content (Harms-Tuohy et al., 2016), the fecal sample, and others. DNA barcoding is also used to identify potential new or cryptic species, as well as to address fundamental environmental and evolutionary issues (Kress & Erickson, 2012). Therefore, DNA barcodes, as a basic screening tool, could provide an insight into the diversity of species, their ecology, biogeography, and evolution (Costa et al., 2018). Identification of species using DNA barcoding methodology begins with the sampling of individuals from different sources (collected on the field, from collections stored in museums, zoos, botanical gardens, seed banks, etc.). Then, DNA is isolated from a small portion of the sample tissue in the laboratory. Through the process of PCR amplification, products are generated which are subsequently sequenced. The resulting sequence is a barcode of the investigated specimen which is then used to search the databases for the match, i.e. the highest probability that the query sequence belongs to some species.

Biodiversity databases play a key role in storing the information on the state of biodiversity which can be used for revitalization and conservation purposes (Turnhout & Boonman-Berson, 2011). Some reference sequence libraries, such as The Spanish Freshwater Fish Database, German Barcode of Life, Norwegian Barcode of Life, Finnish Barcode of Life, Bibliothéque du vivant, Netherlands Barcode of Life, Swiss Barcode of Life, Barcode of Life, and The International Barcode of Life Project are results of projects of European countries implementing high throughput sequencing and DNA barcoding in attempt to capture the genetic diversity (Kalamujić Stroil et al. 2017). One of the most often used global DNA barcode databases is the Barcode of Life Data Systems (BOLD) database which represents a searchable online repository of barcode records, with accompanying data (iBOL, 2017).

### **BARCODE OF LIFE DATABASE SYSTEMS (BOLD) – BOLD baza podataka**

Barcode of Life Database Systems (BOLD) is a web-based and free-of-charge IT system, which has a flexible data model that makes it suitable for projects involving more research teams (Ratnasingham & Hebert, 2007; Kalamujić Stroil et al. 2018b). BOLD consists of four main modules: a data portal, an educational portal, a registry of BINs, and a data collection and analysis workbench (BOLD). BOLD evolved from an informatics workbench for a single, high-volume DNA barcode facility into a resource for the DNA barcoding community. Some features of BOLD are available to any visitor, but system registration provides additional privileges. In addition to the basic package (access to data in public projects and using of the Identification System), registration includes creating private projects and sharing access to password-protected data (Ratnasingham & Hebert, 2007).

#### **The Structure and Content of BOLD Systems**

Four main modules of BOLD are available at <http://www.boldsystems.org/>. Data portal allows for searching over 1.7 M public records using multiple search criteria. It supports queries based on taxonomy, geography, attribution fields, depositories, specimen and sequence identifiers (BOLD). The educational portal, classroom-focused interface to the BOLD database, provides the analysis and publication of DNA barcode data by students. Students can both, explore the large database of DNA barcode records and add their own data to DNA barcode library, all of that monitored by instructors. The Barcode Index Number (BIN) System classifies sequences using well-known algorithms. This page in-

cludes a dendrogram of all member sequences. The workbench is a data storage and analysis platform, supporting the assembly and validation of DNA barcodes and other sequences (BOLD).

### The Statistics Through the Prism of BOLD

To date, over 6,710,000 barcodes have been deposited in BOLD with 196,000 animal species, 68,000 plant species and 22,000 species of fungi and other organism entries (Figure 1).

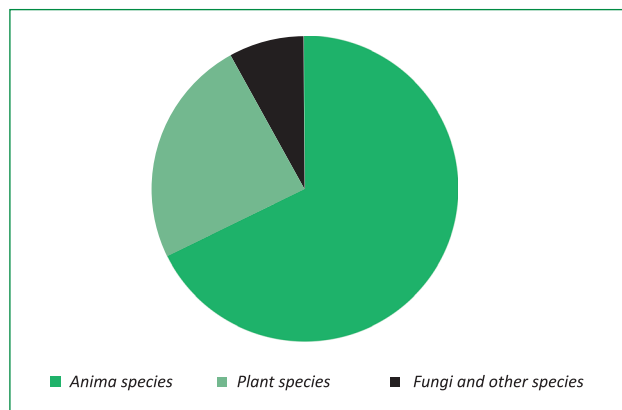


Figure 1. Representation of organisms in the BOLD systems

Grafikon 1. Zastupljenost organizama u BOLD sistemima

BOLD's statistics show very significant data. In order to analyze the involvement of different countries in DNA barcode activities, we chose eight countries: four very developed (Germany, Austria, France, and Sweden) and four countries from the Balkan region (Bosnia and Herzegovina, Croatia, Serbia, and Montenegro). Results showed that Germany has the most public records, with species level sequences and BINs.

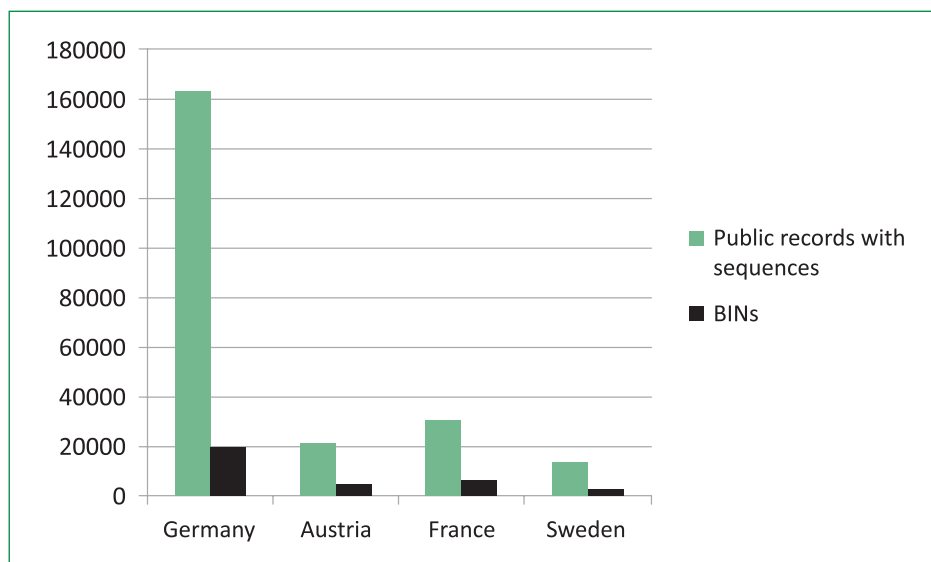


Figure 2. Comparative review of public records with sequences and BINs of four very developed countries

Grafikon 2. Komparativni pregled javnih zapisa sa sekvencama i BIN za četiri vrlo razvijene zemlje

It is followed by France, Austria, and Sweden, respectively (Figure 2). Observing the countries of the Balkan region, Croatia leads in the number of published sequences and BINs (Figure 3).

The top ten orders of sequenced organisms from Bosnia and Herzegovina include Cypriniformes (101), Lepidoptera (80), Entomobryomorpha (49), Decapoda (43), Veneroida (22), Amphipoda (18), Liliales (18), Dip-sacales (14), Salmoniformes (12), and Diptera (11). The highest number of entries are for the following species: *Austropotamobius pallipes* (22), *Congerius kusceri* (14), *Lilium bosniacum* (10), *Gammarus balcanicus* (9), *Troglocaris anophthalmus* (8), *Congerius mulaomerovici* (8), *Rutilus basak* (7), *Chondrostoma phoxinus* (7), and *Telestes dabar* (7).

In this database, there are currently 452 entries for organisms from Bosnia and Herzegovina, which makes 0.0067% of the total number of BOLD entries. According to BOLD statistics, only 1.106% of all organism entries from B&H were submitted by B&H institutions. In comparison to neighboring countries, Serbia and Croatia have deposited 950 (5.26% entries submitted by Serbian institutions) and 2,799 entries (12.08% entries submitted by Croatian institutions), respectively (Figure 4). However, neither of 781 entries for organisms from Montenegro is submitted by Montenegro institutions. For instance, Germany has 162,925 BOLD entries and all of them were submitted by German institutions.

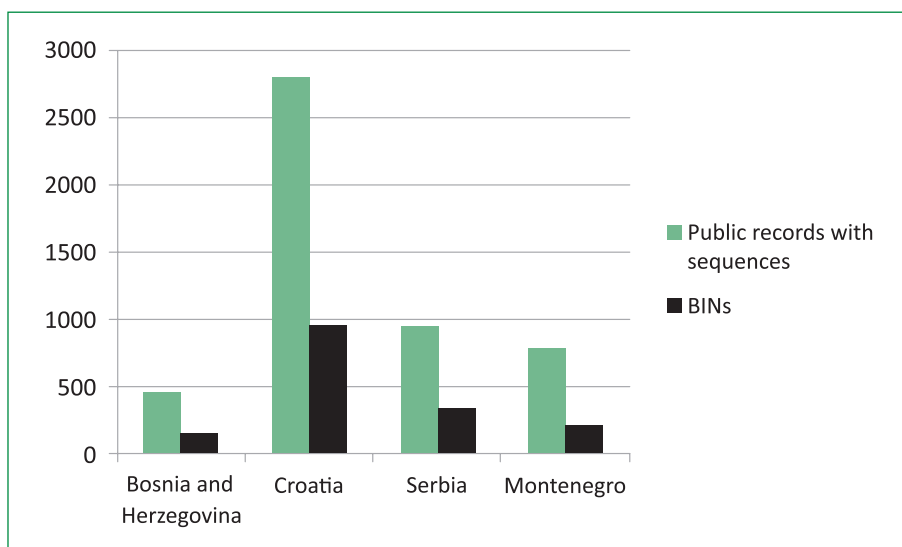
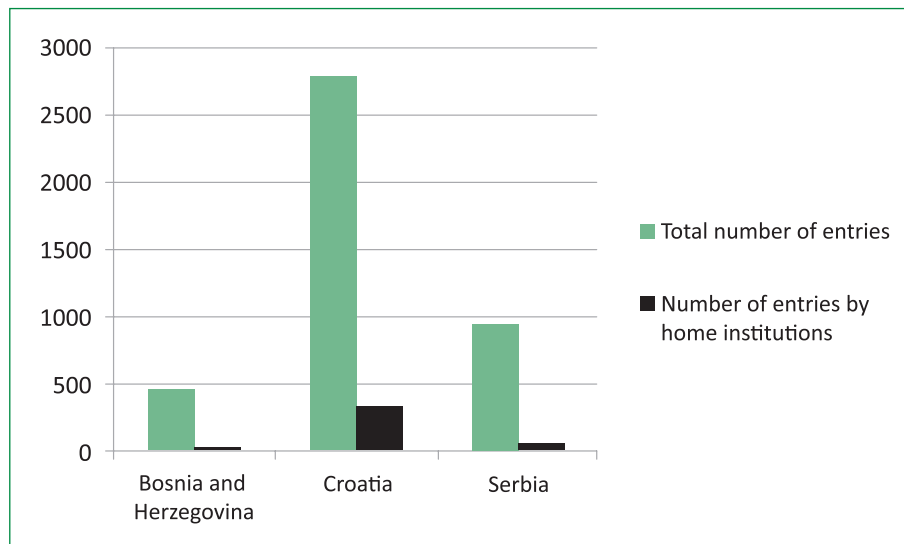


Figure 3. Comparative review of public records with sequences and BINs of the countries of the Balkan region

Grafikon 3. Komparativni pregled javnih zapisa i BIN za zemlje Balkana

Figure 4. Comparative review of total number of entries and number of entries by home institutions

Grafikon 4. Komparativni pregled ukupnog broja podataka i broj podataka unesenih od domaćih institucija



## FUTURE OF DNA BARCODING IN BOSNIA AND HERZEGOVINA – *Budućnost DNK barkodiranja u Bosni i Hercegovini*

Despite the fact that Bosnia and Herzegovina has valuable natural resources with a high percentage of endemic and autochthonous species, BOLD statistics elucidated the lack of coordinated and systematic DNA barcoding research so far. It is necessary to establish continuous progress of molecular-genetic characterization of these resources in the future. It is up to B&H institutions to decide if they want to continue the practice of submitting the data sporadically or if they will animate the research community to actively participate in this global project.

## REFERENCES - *Literatura*

BOLD (2019). Barcode of life. (<http://www.boldsystems.org/index.php>) accessed: 03.02.2019.

Costa, F., Mamos T., Vieira, P.E. (2018). Aquatic macroinvertebrates diversity and evolution: some novel perspectives after 15 years of DNA barcoding [Abstract]. *3rd Central European Symposium for Macroinvertebrate Research*. 19.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.*, 3(5), 294-299.

- Harms-Tuohy, C.A., Schizas, N.V., & Appeldoorn, R.S. (2016). Use of DNA metabarcoding for stomach content analysis in the invasive lionfish *Pterois volitans* in Puerto Rico. *Mar. Ecol. Prog. Ser.*, 558, 181-191.
- Hebert, P.D.N., Cywinska, A., Ball, S.L., & deWaard, J.R. (2003a). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B.*, 270, 313-321.
- Hebert, P.D.N., Ratnasingham, S., deWaard, J.R. (2003b). Barcoding animal life: cytochrome c oxidase subunit I divergences among closely related species. *Proc. R. Soc. Lond. B.*, 270, S96-S99.
- iBOL (2017). What is DNA Barcoding? How DNA Barcoding Works and What it Will Do (<http://www.ibol.org/phase1/about-us/what-is-dna-barcoding/>) downloaded: 03.08.2018.
- Kalamujić Stroil, B., Dorić, S., Hanjalić, J., Lasić, L., Pojskić, N. (2017). Regional biodiversity database (REBIDA) – The first comprehensive database of biological diversity of Bosnia and Herzegovina. *Genetics & Applications*, 1(2), 59-65.
- Kalamujić Stroil, B., Dorić, S., Lukić Bilela, L., Pojskić, N. (2018b). *Applied bioinformatics - Practicum*. Sarajevo: Institute for Genetic Engineering and Biotechnology, Sarajevo.
- Kalamujić Stroil, B., Lasić, L., Hanjalić, J., Mačar, S., & Vesnić, A. (2018a). The first DNA barcode record for *Rhyacophila bosnica* Schmid, 1970 and pairing of adult and larval life stages. *Genetics & Applications*, 2(2), 20-27.
- Kress, W.J., & Erickson, D.L. (2008). DNA barcodes: Genes, genomics, and bioinformatics. *PNAS*. 105, 2761-2762.
- Kress, W.J., & Erickson, D.L. (2012). *DNA barcodes: Methods and protocols*. Springer New York Dordrecht Heidelberg London: Springer Science+Business Media
- Ratnasingham, S., & Hebert, P.D.N. (2007). BOLD: The Barcode of Life Data System. *Mol. Ecol. Notes*. 7(3), 355-364
- Turnhout, E., & Boonman-Berson, S. (2011). Databases, scaling practices, and the globalization of biodiversity. *Ecol. Soc.*, 16, 35. ●

## SAŽETAK

DNK barkodiranje je metoda dizajnirana za pružanje brze i precizne identifikacije vrsta korištenjem jedne ili više kratkih sekvenci gena koje se nazivaju barkodovi. U većini studija biljaka i gljivica standardni odabrani crtični kodovi su tri plastidna (*rbcl*, *matK* i *trnH-psbA*) i jedan nuklearni (*ITS*) genski region. Relativno visoka i očuvana stopa evolucije mtDNA u sekvencama učinila je *COI* markerom izbora kod životinja. BOLD je besplatno dostupna platforma za skladištenje i analizu podataka zasnovana na oblaku, a razvijena s ciljem unapređenja nauke o biološkoj raznolikosti kroz identifikaciju vrsta s barkodom DNK. Do danas je preko 6 miliona barkodova deponovano u BOLD sa 196 000 životinjskih vrsta, 68 000 biljnih vrsta i 22 000 vrsta gljivica i drugih unosa u organizme. U ovoj bazi podataka trenutno postoji 447 unosa za organizme iz Bosne i Hercegovine, što čini 0,0067% od ukupnog broja BOLD unosa. Prema BOLD statistikama, samo 1,11% svih unosa organizama iz BiH su podnijele institucije BiH.

Uprkos činjenici da Bosna i Hercegovina ima vrijedne prirodne resurse s visokim procentom endemskih i autohtonih vrsta, BOLD statistika je rasvijetlila nedostatak koordiniranog i sistematskog istraživanja barkodiranja DNK do sada. Potrebno je uspostaviti kontinuirani napredak molekularno-genetske karakterizacije ovih resursa u budućnosti. Na bh. Institucijama je da odluče žele li nastaviti s praksom povremenog dostavljanja podataka ili će animirati istraživačku zajednicu da aktivno učestvuje u ovom globalnom projektu.