https://doi.org/10.52388/1857-064X.2022.3.08

IN SILICO VALIDATED PRIMERS FOR METABARCODING TERRESTRIAL VERTEBRATE SPECIES FROM THE REPUBLIC OF MOLDOVA

Victor Sitnic

Institute of Zoology

Rezumat

În studiul de față ne-am propus design-ul și testarea *in silico* a unor primeri care ar putea fi utilizați pentru studii de metabarcodare a unor specii de vertebrate terestre din Republica Moldova. Utilizând tehnici bioinformatice, au fost determinate acoperirea taxonomică, specificitatea și gradul de conservare a noilor primeri în comparație cu primerii universali utilizați pentru metabarcodarea vertebratelor. Grupul țintă a fost format din 174 specii de vertebrate terestre care fac parte din fauna Republicii Moldova și au mitogenomul inclus în baza de date globală *RefSeq*. Noii primeri au demonstrat performanțe superioare în ce privește grupul de specii studiat.

Cuvinte cheie: primeri, metabarcodare, ADN de mediu, in silico, vertebrate terestre.

Depus la redacție: 24 noiembrie 2022

Adresa pentru corespondență: Sîtnic Victor, Institutul de Zoologie, , str. Academiei 1, MD-2028 Chișinău, Republica Moldova, e-mail: victor.sitnic@zoology.md

Introduction

Metabarcoding of environmental DNA (eDNA) is a relatively new biomonitoring method and refers to the molecular analysis of a mixture of intra- and extracellular DNA from various organisms. This allows the simultaneous characterization and identification of more than one taxon using a sequence of a genomic region that is suitable for this purpose. The eDNA analysis of vertebrate species is an under development tool for studying ecosystem biodiversity and like any metabarcoding research requires a correct selection of primers. Monitoring the diversity of terrestrial vertebrates using only traditional methods is generally laborious and difficult, thus currently there are increasingly pronounced trends which follow the development of eDNA techniques for terrestrial vertebrate research. A recent study carried out in the Copenhagen zoo allowed the identification of 49 vertebrate species (30 mammal, 13 bird, 4 fish, 1 amphibian, and 1 reptile species) based on airborne eDNA [5]. In their research they used the universal vertebrate primers *12S-V5* [15] and the pair *16Smam* designed to amplify mammals [13]. Another metabarcoding research aimed to compare the use of camera trapping with isolation and analysis of mammals eDNA from soil proved that the last method was more effective, in particular for small species, which do not trigger camera traps [10]. There are also recent studies aimed to use public databases and bioinformatics algorithms for the purpose of *in silico* validation of different primers [6, 11]. In our study we aimed to design and perform *in silico* testing of a pair of primers that could be used for

metabarcoding studies of terrestrial vertebrates from the Republic of Moldova. Also, using electronic PCR software, we set out to determine its taxonomic coverage (Bc), specificity (Bs) and degree of conservation in comparison with other universal primers used for vertebrate studies: *12S-V5-F* and *12S-V5-R*. The pair *12S-V5* targets the mitochondrial 12S rRNA region and was designed by Tiayyba Riaz and colleagues in 2011 [15]. It was validated using both: bioinformatic algorithms and laboratory experimental approaches [15] and it shows a good performance compared to other universal primers used for the identification of vertebrate species [9,18].

Metabarcoding primers are used to amplify DNA from a group of more or less taxonomically related organisms that have conserved sequences for binding these primers. Also it is necessary that the amplified metabarcodes to contain variable regions for different species, this being important for taxonomic identification. Ideally, the primers should have high specificity for the target species group and broad taxonomic coverage within the group. This allows a comprehensive and precise identification of the studied species. In the present research, the target group consisted of terrestrial vertebrates species that are part of the fauna of the Republic of Moldova and have their mitogenomes included in the global *RefSeq* database [12]. The group includes 127 species of birds, 38 species of mammals, 5 species of amphibians and 4 species of reptiles. *RefSeq* sequences are of high quality and contribute to a good metabarcoding research.

Materials and methods

In order to perform the research there were used *ecoPrimers* [15], *ecoPCR* [7] and *OBITools* [4] bioinformatics instruments. These programs are used for the analysis of genomic data in metabarcoding studies and are distributed as an open source software. The algorithm *ecoPrimers* can be utilized for identification of new metabarcode markers and their associated primers while *ecoPCR* allows to test them by performing *in silico* PCR using a database of sequences (associated with species from the target group). Together, these programs, allow to design and conduct *in silico* validation of PCR metabarcoding primers. The online oligo analysis tool *PrimerROC* [8] was additionally used for primers characterization. In order to simulate real PCR conditions, it is necessary that the *in silico* PCR algorithm to allow some mismatches between primers and their target sequences. Since two mismatches represent a reasonable value that can ensure a qualitative amplification of the studied eDNA it has been chosen a maximum of two errors per primer. More than three mismatches for any of the primers may cause an inefficient amplification. Also, to ensure a good quality of the amplification process, no mismatches were allowed for the last 2 nucleotides from the 3' flank of each primer. Mismatches in these regions could prevent primer alignment and strongly affect the efficiency of the amplification process. The assessment of primers conservation and graphical representation were performed using *R language* [14] and the *ROBITaxonomy* [1], *ROBITools* [2] and *ROBIBarcodes* [3] libraries. The reference mitogenomes database was created using the NCBI platform [16] and contains 174 reference mitochondrial genomes (one genome per species). This dataset was derived from the individual verification in *RefSeq* of 337 terrestrial vertebrate species from Moldova's fauna.

Results and discussions

Following the run of the *ecoPrimers* algorithm on the 174 *RefSeq* mitogenomes there were identified several primer pairs with a size of 18 base pairs and an associated metabarcode length between 80 and 150 bp. Pairs with a melting temperature difference between forward and reverse primers greater than 3°C were ignored. It has been selected a pair with an optimal dG value (-0,83 kcal/mol) and with melting temperature of 53,9 °C for the forward primer (AAGGCGGATTTAGCAGTA) and 51,0 °C for the reverse (CACTTACCTTGTTACGAC). As far as the pair amplify a region of 12S rRNA mitochondrial gene and was designed for metabarcoding research it was provisionally named *Met-12S* (respectively *Met-12S-F* for the forward primer and *Met-12S-R* for the reverse). *In silico* amplification of the 174 sequences with the *12S-V5* and *Met-12S* primers revealed that the Riaz's pair has a higher degree of conservation compared to new primers. In the figures 1 and 2 are shown the sequence logos resulted from the analysis of the two pairs. Sequence logos are graphical representations that reflect efficiently the conservation of forward and reverse primers for a given group of sequences. The letters represent deoxyribonucleotide abbreviations, and the height of the letters is an indication of the degree of conservation at different positions (the height of the letters is directly proportional to the frequency of the complementary nucleotides in the associated positions) [17].

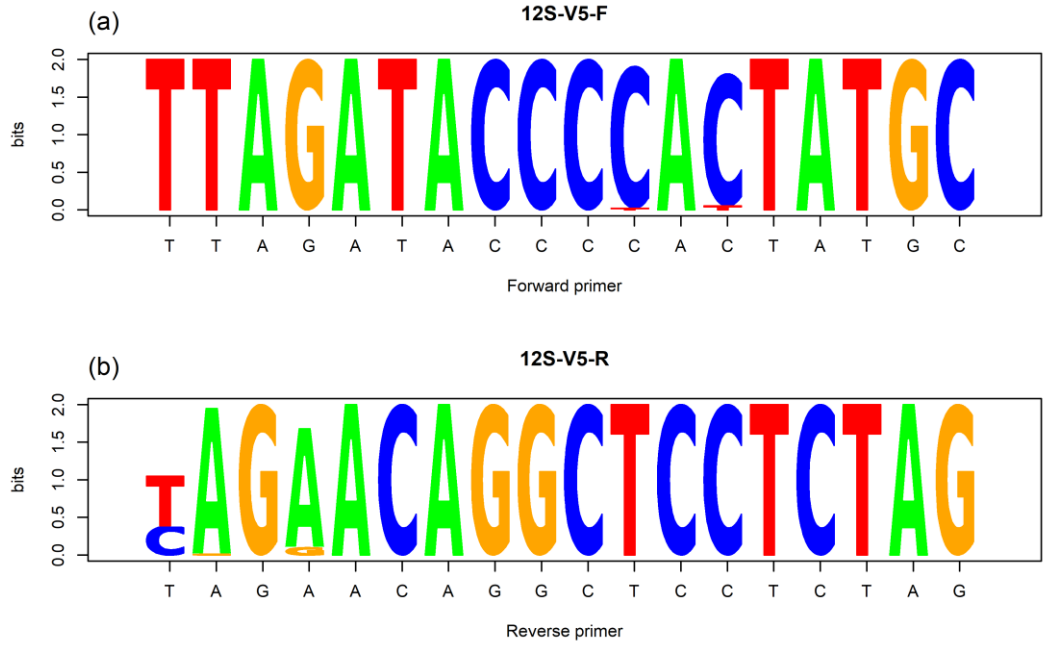

Figure 1. Sequence logos obtained for primer pair 12S-V5:
a – forward primer (12S-V5-F) and b – reverse primer (12S-V5-R)

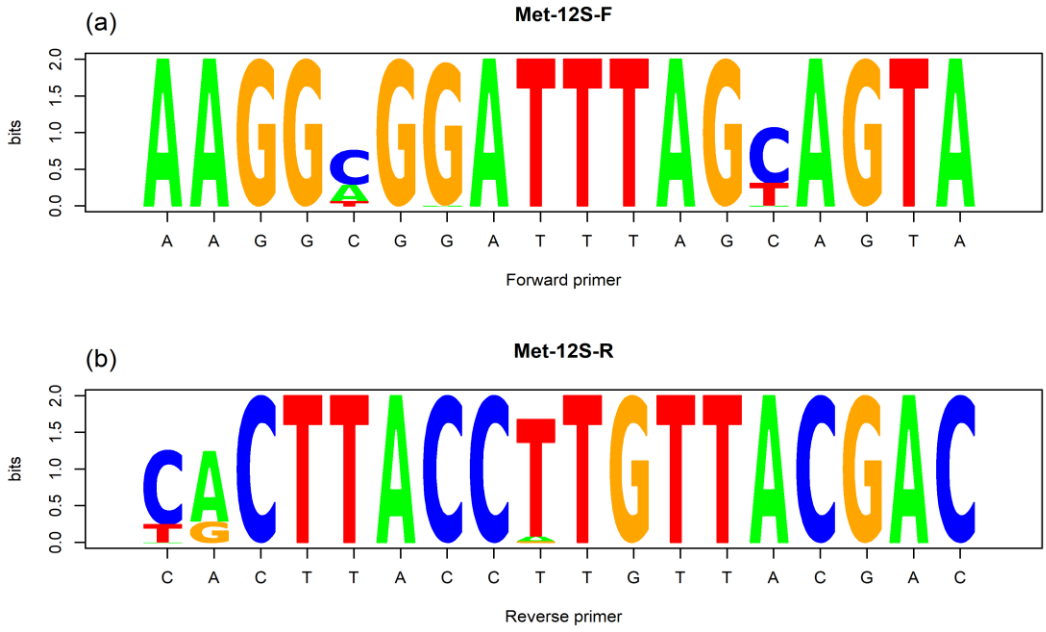

Figure 2. Sequence logos obtained for primer pair Met-12S:
a – forward primer (Met-12S-F) and b – reverse primer (Met-12S-R)

Thus, it can be seen that at the 3' flank all the primers have a high degree of conservation. A lower degree may be observed in the first position of 5' flanking region of *12S-V5-R* and also in positions 1, 2 of the *Met-12S-R* 5' flank. The number of mismatches per primer is shown in Figure 3.

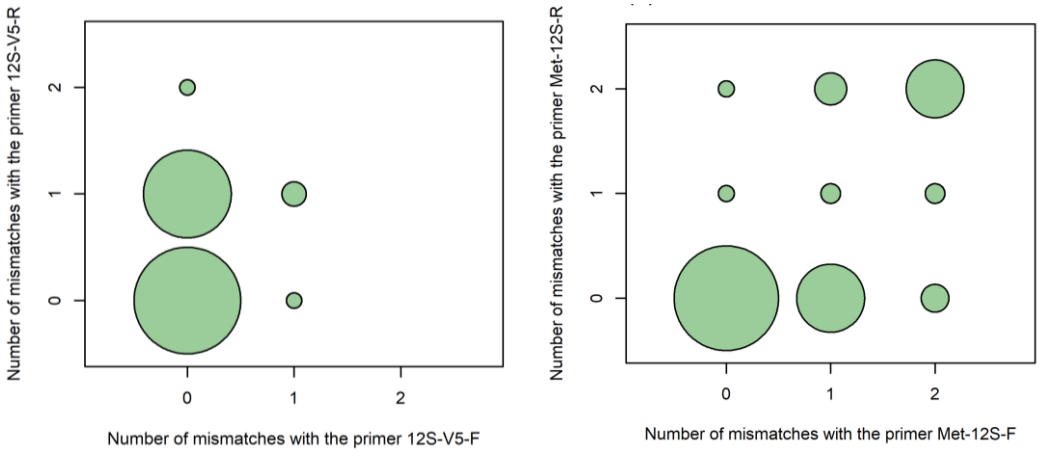

Figure 3. Number of mismatches for 12S-V5 (left) and Met-12S (right) primer pairs

It can be noticed that the new pair presents a maximum number of 2 mismatches for both primers while *12S-V5* - a maximum of one mismatch for the forward primer and two for the reverse. The number of electronically amplified species can be increased by

allowing more than two mismatches per primer however this may result in a failed *in vitro* PCR reaction.

The primers performance was evaluated in the basis of two indexes:

1. ***Bc*** (Coverage Index) – corresponds to the ratio between the number of amplified taxa and the total number of taxa in the reference database [7].
2. ***Bs*** (Specificity Index) – is defined as the ratio between the number of correctly identified taxa and the number of amplified taxa [7].

These two indexes demonstrated a superior performance for pair *Met-12S* having a species level *Bc* value of 0.9828 (98.28%) and *Bs* – 0.8655 (86.55%). Respectively, for the *12S-V5* pair resulted a *Bc* of 0.9770 (97.70%) and *Bs* – 0.5353 (53.53%). Thus, according to the bioinformatics algorithm, the *Met-12S* pair presents taxonomic coverage for 171 species out of 174 and correctly identifies 148 out of 171, while Riaz's universal primers amplify 170 out of 174 and identify 91 out of 170. In the end it is necessary that the *Bc* and *Bs* values to be as high as possible and the maximum and average length of the metabarcodes – as small as possible. As environmental DNA is usually highly degraded, it is recommended that metabarcodes to be less than 150 bp in length. This is important especially for monitoring species diversity when using such samples as faeces or soil. The metabarcode lenght associated with new designed primers ranges from 117 to 136 bp (Figure 4).

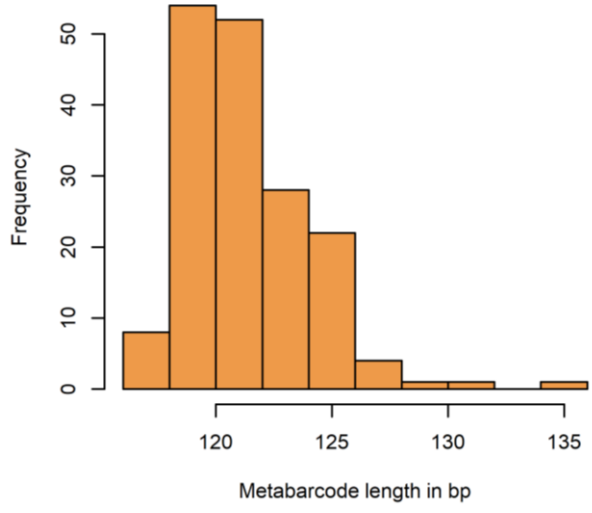

Figure 4. Metabarcode length of Met-12S primer pair

It is necessary to mention that in public databases may be found multiple sequencing errors as well as other ambiguities and in order to obtain a good estimation of primers coverage and resolution power is required a careful selection of reference sequences (eg: sequences from the *RefSeq* database). The above presented results were obtained on the basis of *in silico* research using 174 high quality sequences available in the *RefSeq* database. Also, we created a local reference database representative of these sequences which could be used in future metabarcoding studies and corresponds to 174 species, 130 genera, 61 families, 26 orders and 4 classes. As can be observed most species are

the unique representative of their genus. The local database contains one sequence per species this having a role of reducing aspects related to overrepresentation.

Although effective in monitoring biodiversity, environmental DNA research methods are not widely applied in the Republic of Moldova and are still in their early stage. The application of bioinformatics methods in conjunction with the establishment of laboratory conditions will enable the implementation of sequencing protocols and the use of environmental DNA as a monitoring tool for the local species.

The research was performed within the institutional project no. 20.80009.7007.02 *Evolutionary changes of economically important terrestrial fauna, of rare and protected species in the conditions of anthropogenic and climatic changes* – EVOLANTER.

Conclusions

The bioinformatics research of the studied vertebrate mitogenomes allowed the *in silico* validation of a pair of candidate primers that targets 174 terrestrial vertebrate species from the Republic of Moldova. According to performed *in silico* research, the new designed primers provisionally named *Met-12S*, demonstrated a superior performance in comparison with universal vertebrate primers *12S-V5*. However, there would be a need for wet lab testing to further validate its metabarcoding efficacy.

Aknowledgement. The research was performed within the national project no. 20.80009.7007.02 „Evolutionary changes of economically important terrestrial fauna, of rare and protected species in the conditions of anthropogenic and climatic changes", State Programe (2020-2023).

References

1. alpine L-Ld (2012). _ROBITaxonomy: Metabarcoding data biodiversity analysis_., 2012, R package version 0.1.

2. alpine L-Ld (2012). _ROBITools: Metabarcoding data biodiversity analysis_. R package version 0.1.

3. alpine L-Ld (2013). _ROBIBarcodes: Metabarcoding barcode database_. R package version 0.1.

4. *Boyer F., Mercier C., Bonin A., Taberlet P., Coissac E.*, OBITools: a Unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 2014, submitted.

5. *Casper V. J., Lynggaard C., Bertelsen M. F., Johnson M. S., Tobias G. F., Morten T. O., Bohmann K.*, Airborne environmental DNA for terrestrial vertebrate community monitoring, Current Biology, Volume 32, Issue 3, 2022, Pages 701-707.e5, ISSN 0960-9822, https://doi.org/10.1016/j.cub.2021.12.014.

6. *Chukwuemeka P.O., Umar H.I., Olukunle O.F., Oretade O.M., Olowosoke C.B., Akinsola E.O., Elabiyi M.O., Kurmi U.G., Eigbe J.O., Oyelere B.R., Isunu L.E., Oretade O.J.*, In silico design and validation of a highly degenerate primer pair: a systematic approach. J Genet Eng Biotechnol., 2020 Nov 17;18(1):72. doi: 10.1186/s43141-020-00086-y. PMID: 33205353; PMCID: PMC7671957.

7. *Ficetola G.F., Coissac E., Zundel S., Riaz T., Shehzad W., Bessière J., Taberlet P., Pompanon F.*, An in silico approach for the evaluation of DNA barcodes. BMC Genomics., 2010 Jul 16;11:434. doi: 10.1186/1471-2164-11-434. PMID: 20637073; PMCID: PMC3091633.

8. *Johnston A. Lu. J. et al.*, PrimerROC: accurate condition-independent dimer prediction using ROC analysis, 2018, Sci. Rep. http://www.primer-dimer.com/ .

9. *Kitano T., Umetsu K., Tian W., Osawa M.*, Two universal primer sets for species identification among vertebrates. Int J Legal Med., 2007 Sep;121(5):423-7. doi: 10.1007/s00414-006-0113-y. Epub 2006 Jul 15. PMID: 16845543.

10. *Leempoel K., Hebert T., Hadly E.A.*, A comparison of eDNA to camera trapping for assessment of terrestrial mammal diversity. Proceedings of the Royal Society B: Biological Sciences, 2020, 287: 20192353.

11. *Licuanan A, Matias A.M.*, In silico evaluation of the taxonomic resolution and coverage of the COI region and alternative barcode markers for Echinoderms. Philipp J Sci, 2022, 151(3): 955–968. https://doi.org/10.56899/151.03.14.

12. *O'Leary N.A., Wright M.W., Brister JR., Ciufo S., Haddad D., McVeigh R., Rajput B., Robbertse B., Smith-White B., Ako-Adjei D., Astashyn A., Badretdin A., Bao Y., Blinkova O., Brover V., Chetvernin V., Choi J., Cox E., Ermolaeva O., Farrell C.M., Goldfarb T., Gupta T., Haft D., Hatcher E., Hlavina W., Joardar V.S., Kodali V.K,. Li W., Maglott D., Masterson P., McGarvey K.M., Murphy M.R., O'Neill K., Pujar S., Rangwala S.H., Rausch D., Riddick L.D., Schoch C., Shkeda A., Storz S.S., Sun H., Thibaud-Nissen F., Tolstoy I., Tully R.E., Vatsan A.R., Wallin C., Webb D., Wu W., Landrum M.J., Kimchi A., Tatusova T., DiCuccio M., Kitts P., Murphy T.D., Pruitt K.D.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res, 2016 Jan 4;44(D1):D733-45 PubMed

13. *P-Taylor G.*, Reproducibility of ancient DNA sequences from extinct Pleistocene fauna., Molecular Biology and Evolution, Volume 13, Issue 1, Jan 1996, Pages 283–285, https://doi.org/10.1093/oxfordjournals.molbev.a025566.

14. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2021, Vienna, Austria. URL https://www.R-project.org/.

15. *Riaz T., Shehzad W., Viari A., Pompanon F., Taberlet P., Coissac E.*, ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Res. 2011 Nov;39(21):e145. doi: 10.1093/nar/gkr732. Epub 2011 Sep 19. PMID: 21930509; PMCID: PMC3241669.

16. *Sayers E.W., Bolton E.E., Brister J.R., Canese K., Chan J., Comeau D.C., Connor R., Funk K., Kelly C., Kim S., Madej T., Marchler-Bauer A., Lanczycki C., Lathrop S., Lu Z., Thibaud-Nissen F., Murphy T., Phan L., Skripchenko Y., Tse T., Wang J., Williams R., Trawick B.W., Pruitt K.D., Sherry S.T.*, Database resources of the national center for biotechnology information. Nucleic Acids Res., 2022 Jan 7;50(D1):D20-D26. doi: 10.1093/nar/gkab1112. PMID: 34850941; PMCID: PMC8728269.

17. *Schneider T.D., Stephens R.M.*, Sequence logos: a new way to display consensus sequences. Nucleic Acids Res., 1990 Oct 25;18(20):6097-100. doi: 10.1093/nar/18.20.6097. PMID: 2172928; PMCID: PMC332411.

18. *Sitnic V.*, Comparative in silico analysis of primers used for vertebrate metabarcoding. In: Life sciences in the dialogue of generations: connections between universities, academia and business community. R, September 29-30, 2022, Chisinau, Republic of Moldova: Moldova State University, 2022, p. 144.