



Fish DNA Barcoding: A Comprehensive Survey of Bioinformatics Tools and Databases

14

Rupali C. Mane, Ganesh Hegde, Ravi Prabhakar More, Rajesh Ramavadh Pal, and Hemant J. Purohit

Abstract

A paradigm shift took place with the advent of molecular taxonomy, which is a combinatorial approach utilizing both computational and molecular biology. DNA barcoding is a reliable, cost-effective method that uses the cytochrome *c* oxidase I (COI) mitochondrial gene to recognize animal species. This gene has a short subsequence 658 bp region that is used for species discrimination. The availability of amplification standard operation protocols and sequence databases for barcoding enables the use of COI sequences for studying taxonomic aspects, particularly in phylogeny, phylogeography, and population genetics studies. The overall process of DNA barcoding in fish is widely performed under the umbrella of molecular and computational methods. In this chapter, we report the current status of fish DNA barcoding with respect to the databases and software tools available in the public domain.

R. C. Mane

Department of Oral and Maxillofacial Surgery, Faculty of Dental Sciences, M.S. Ramaiah University of Applied Sciences, Bangalore, Karnataka, India

G. Hegde

Central Institute of Freshwater Aquaculture-ICAR (CIFA), Regional Research Center, Bangalore, Karnataka, India

R. P. More (✉)

ADBS, TIFR-National Centre for Biological Sciences (NCBS), Bangalore, Karnataka, India
e-mail: ravimore7@yahoo.in

R. R. Pal

Nagarjuna Fertilizers and Chemicals Limited, Hyderabad, Telangana, India

H. J. Purohit (✉)

Environmental Biotechnology and Genomics Division, CSIR-National Environmental Engineering Research Institute (NEERI), Nagpur, Maharashtra, India
e-mail: hj_purohit@neeri.res.in; hemantdrd@hotmail.com

Keywords

Fish DNA barcoding · Cytochrome *c* oxidase I (COI) mitochondrial gene · Barcode of Life Data System (BOLD) · FISH-BOL · DNA barcoding tools and databases

14.1 Introduction

Bioinformatics has emerged into a fully fledged multidisciplinary field that integrates statistics and informatics for the analysis of biological data. Due to the advancement in next-generation sequencing (NGS) technology, there has been a dramatic growth in studies of fish genomics (Kumar and Kocour 2017). Public databases now host a catalogue of complete genomes of biological species (mainly fish), which contain protein sequences, protein three-dimensional structures, metabolic pathways, and biodiversity-related information (Vera-Escalona et al. 2017; Adrian-Kalchhauser et al. 2017). Bioinformatics is helping to solve biological problems using software and databases in areas such as functional genomics, bimolecular structure, proteome analysis, taxonomy, and pesticide molecule design (Cambiaghi et al. 2016).

Our earth harbors approximately 8.7 million species, of which around 2.2 million are marine (Mora et al. 2011). IUCN Red List version 2016–3 estimates that the number of described fish species is 33,400. The challenging part was to identify and classify this many species. Earlier methods employed to identify species relied mainly on morphology, protein electrophoresis, and chromatography (Yilmaz et al. 2007; Strauss and Bond 1990; Viswanathan and Pillai 1956). The barcoding technique is effectively utilized in fisheries and has been used to identify recently radiated megadiverse fauna from neotropical areas. The mitochondrial gene encoding cytochrome *c* oxidase subunit I (COI) is used as a marker in phylogeny, phylogeography, and population genetics studies (Pereira et al. 2012; Sbordoni 2010). It has been used for systematic study of native freshwater fish, to monitor the geographic distribution of species (Hubert et al. 2008), and to monitor threatened shark species (Velez-zuazo et al. 2015). These applications facilitate authentication of commercially important species and thereby enhance transparency and fair trade in the domestic fisheries market (Cawthorn et al. 2012). Recent developments include meta-barcoding, in which DNA released by organisms into the environment (eDNA) via cells, excreta, gametes, and decaying materials can effectively be used for species identification. A study conducted in the English Lake District described fish communities in large lakes, both quantitatively and qualitatively (Hanfling et al. 2016). The DNA meta-barcoding approach is considered a next-generation tool for biodiversity monitoring in aquatic ecosystems (Valentini et al. 2016). Mini-barcode primer pairs of length 127–314 bp were developed for authentication of fish food products (Shokralla et al. 2015).

In 2004, an international initiative by the Consortium for the Barcode of Life (CBOL) was taken to make DNA barcoding a standard method or tool for

identification of species (<http://www.barcodeoflife.org/content/about/what-cbol>) (Group et al. 2009). The Barcode of Life Data System (BOLD) is the central informatics platform for DNA barcoding (ibol.org). The Fish Barcode of Life (FISH-BOL) and Shark Barcode of Life (Shark-BOL) initiatives are two important fish barcoding projects at the global level. In India, the Fish Barcode Information System (FBIS), a DNA barcode database on fish, was developed by the National Bureau of Fish Genetic Resources (NBFGR). The overall process of DNA barcoding in fish exploits both molecular and computational methods. A unique region of the specimen is considered as a barcoding marker. In the case of fish, the marker is the gene encoding cytochrome *c* oxidase I (COI) (Hebert et al. 2003).

The general strategy of barcoding involves DNA extraction from the specimen, amplification of a unique marker region using the polymerase chain reaction (PCR), and sequencing. Computational aspects such as editing and aligning sequences is carried out using software such as BOLD v 3.0 (Pereira et al. 2012), TaxI (Steinke et al. 2005), MEGA (Kumar et al. 2008), MEGA 5.05 (Landi et al. 2014), CodonCode Aligner 3.7.1.1 (Shokralla et al. 2015), and GENIOUS PRO 5.4.2, (Henriques et al. 2015). Results are later submitted to GenBank or BOLD databases. Hence, once sequencing is completed, the computational aspect plays a key role not only in identification but also in addressing questions related to evolution, diversity (Shen et al. 2016), and taxonomy (Hebert and Gregory 2005).

14.2 Molecular and Computational Approaches for Fish DNA Barcoding

The tissue sample collected from the fish specimen is subjected to DNA extraction. PCR amplifies the target COI gene using a universal primer cocktail (Ivanova et al. 2007). Sequencing of amplified PCR products by BigDye Terminator v.3.1 Cycle Sequencing Kit (Cawthorn et al. 2012) gives both forward and reverse strand sequences. Subsequent important steps are editing, alignment, and sequence submission.

A full-length sequence is made up of aligned reverse and forward strand sequences for all samples of a species (<http://mail.nbfgr.res.in/fbis/protocol.php>). All the aligned sequences are translated into amino acids to approve the efficiency of the sequence and to identify the presence of nuclear DNA pseudogenes, insertions, deletions, or stop codons (Shen et al. 2016). Edited sequences are placed into the BLAST tool of the National Center for Biotechnology Information (NCBI) to obtain the nearest similar sequence matches and are later submitted to GenBank or BOLD. (<http://mail.nbfgr.res.in/fbis/protocol.php>). Available editing packages are DNASTAR multiple packages (Chen et al. 2015), Sequencer 4.8 (Gene Codes) (Velez-zuazo et al. 2015), GAP 4 (Shirak et al. 2016; Baxevanis and Ouellette 2004), MEGA version 4.1 (Costa et al. 2012), and MEGA 5.05 (Landi et al. 2014). Useful software packages, alignment tools, databases, and web pages pertaining to barcoding and other related analysis are listed in Tables 14.1, 14.2.

Table 14.1 Fish DNA barcoding databases

| Database for barcoding | Website |
|------------------------|---|
| BOLD | http://www.boldsystems.org/ |
| FBIS | http://mail.nbfgres.in/fbis/protocol.php |
| FISH-BOL | http://www.fishbol.org/ |
| iBOL | http://www.barcodeoflife.org/ |
| NCBI GenBank | https://www.ncbi.nlm.nih.gov/genbank/barcode/ |

Table 14.2 Software used for DNA barcoding

| Software | Type | Website |
|--------------|----------------------------|---|
| BioEdit | Alignment | www.mbio.ncsu.edu/bioedit/bioedit.html |
| MUSCLE | Alignment | www.ebi.ac.uk/Tools/msa/muscle/ |
| CLUSTALW2 | Alignment | www.ebi.ac.uk/Tools/msa/clustalw2/ |
| GENEIOUS | Alignment | www.geneious.com/ |
| CLC Genomics | Alignment | https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/ |
| PHYLIP | Phylogenetic | evolution.genetics.washington.edu/phylip.html |
| MrBayes | Phylogenetic | mrbayes.sourceforge.net/ |
| DNASTAR | Alignment/ phylogenetic | https://www.dnastar.com/ |
| MEGA | Alignment/ phylogenetic | www.megasoftware.net/ |

Sequence alignment is a method for finding commonality and conserved sequence regions between two or more sequences using a statistical algorithm. It is an important step in identifying the functional, structural, and evolutionary roles of a molecular sequence. A number of sequence alignment packages are available, among which BLAST (Altschul et al. 1990; Madden 2013), MUSCLE (Henriques et al. 2015), CLUSTALX 2.0 (Chen et al. 2015), ClustalW (Velez-zuazo et al. 2015), SeqScape v. 2.1.1 (Applied Biosystems. Inc.) (Zhang and Hanner 2012), BOLD v.3.0 (Pereira et al. 2012), and CodonCode Aligner v 3.7.1.1 (CodonCode Corp., Dedham, MA, USA) (Shokralla et al. 2015) are routinely used.

The usefulness of DNA barcode data in deciphering the phylogenetic relationship between and within species is well studied and involves a series of steps such as alignment, determination of substitution model, and tree building. The latter includes either distance-based tree building or character-based tree building. The distance-based method utilizes the distance between two aligned sequences to generate phylogenetic trees, whereas character-based methods use the composition of oligonucleotide frequencies (e.g., di-, tri-, tetra-, penta-, hexa-, heptanucleotides) in the sequences (Baxeavanis and Ouellette 2004; Higgs and Manchester 2001). The most commonly employed distance-based methods are neighbor-joining (Saitou and Nei 1987), the Fitch–Margoliash method, the unweighted pair group method with arithmetic mean (UPGMA), and minimum evolution (ME). Maximum parsimony (MP) and maximum likelihood (ML) are two major character-based methods

used for phylogenetics (Felsenstein 1981). In addition, Bayesian analysis has been proposed for phylogeny (Huelsenbeck and Ronquist 2001). Tests for evaluating constructed trees include the skewness test, permutation test, and bootstrapping, which can be parametric or nonparametric, and the likelihood ratio test. Software packages for phylogenetic analysis include PHYLIP, PAUP, PUZZLE, FastDNAm1, MACCLADE, and MOLPHY, along with internet-accessible phylogenetic software such as WEBPHYLIP, PhyloBLAST, BLAST 2, and Orthologue Search Server (Baxevanis and Ouellette 2004).

Noncoding internal transcribed spacer genes have also been suggested as candidate barcodes, along with the COI gene for animal and plant DNA barcoding (Gao et al. 2017; Yang et al. 2017). Two new approaches (DV-RBF and FJ-RBF) have been used to align the noncoding regions for DNA barcoding and showed 100% success rate in identifying marine fish species. (Zhang et al. 2012). On other hand, alignment-free methods such as normalized compression distance (NCD) and information-based distance (IBD) have been utilized for taxonomic analysis of barcode sequences (La Rosa et al. 2013). Taxonomic classification methods are mainly categorized into (1) tree-based approaches, (2) composition-based approaches, (3) similarity-based approaches, and (4) hybrids. These methods required reference databases to predict the taxonomy (Tanabe and Toju 2013).

In a recent study, similarity-based methods such as nearest-neighbor, centric auto-k-NN (NN Cauto), and query-centric auto-k-NN (Q Cauto) were proposed for barcoding studies (Tanabe and Toju 2013). A method of string kernel-based sequence analysis of barcode data sets was proposed that considerably improves species identification accuracy compared with traditional approaches (Kuksa and Pavlovic 2007). The few sequence identification methods that use pairwise alignment (e.g., BLAST) are not able to discriminate species that have highly similar sequences, because only very few base pairs are different between the sequences. To address this issue, alignment-free methods (e.g., BRONX) were developed to identify species sequences (Little 2011). BRONX detects short subsequence regions and matching regions in reference sequences. Based on these regions, the algorithm generates a score without use of multiple sequence alignment to identify sequences at the genus level (Little 2011).

14.3 Public Domain Databases

Recent progress in next-generation sequencing (NGS) platforms has led to advancement of the discipline of bioinformatics for the annotation of genome data. Public databases contain huge amounts of accessible data on whole genome sequences, which have improved research in applied fish science. There are some very popular primary, secondary, and specialized databases available from BOLD, FISH-BOL, GenBank, and FBIS.

14.3.1 Barcode of Life Data System

The Barcode of Life Data System (BOLD) (<http://www.barcodinglife.org>) facilitates a detailed collection of specimens deposited by researchers from different barcoding studies. This database holds three main categories of information. The first category is basic information on the specimen and sequence entries. The second maintains quality assurance and manages barcode data with all related information. The third category facilitates a detailed catalogue of specimen data entries from geographically different researchers. A user can store specimen information in the following sections:

- Species name
- Voucher data, institution storing, and catalogue number
- Collection record, which includes collector name, location with GPS coordinates, and data of collection
- Identifier of the specimen
- COI sequence with minimum 500 bp
- PCR primers referred for amplicon capture of trace files

BOLD is an informatics workbench used for collection, storing, scrutiny, and publication of DNA barcode entries and is freely accessible. It involves more than 65,000 lines of combined code written in Java, C++, and PHP. To gain formal barcode status, certain criteria must be satisfied, including species name, voucher data, and collection record. BOLD employs many tools to identify data anomalies or low-quality records. All acquiesced sequences are translated into amino acids and are matched against a hidden Markov model (HMM) of COI protein to confirm that they essentially originate from the COI sequence. Later sequences are checked for stop codons, and also against a small set of possible contaminants. If any errors are detected, the submitter is informed and the sequence is flagged. After providing a trace file, BOLD further determines a PHRED score for each nucleotide position and a mean value for the full sequence based on these results. Next, it manages each sequence entry into one of four classes: failed (no sequence), low quality (mean PHRED < 30), medium quality (mean PHRED = 30–40), and high quality (mean PHRED > 40). The data stored in BOLD can be readily exported in FASTA format for use in other analytical packages. BOLD provides an examination utility that permits users to determine sequence coverage for a specific taxonomic or geographic region. It includes an integrated analytic system (MAS), which provides data analysis tools such as the taxon identification (ID) tree. Unknown sequences are identified by pasting their sequence record into the input box on the ID form. Core data element records in BOLD consist of a specimen page and a sequence page. Barcodes in the search archives are grouped into two categories. Species are considered with three representatives and maximum divergence of 2%. A HMM method is used to align the query sequence with archive sequences. The HMM method is faster than BLAST because of its efficient data processing capability.

BOLD detects species if the query sequence displays a close match with at least <1% divergence against the archive sequences (Ratnasingham and Hebert 2007).

14.3.2 Fish Barcode of Life Campaign and Fish Barcode Information System

The campaign FISH-BOL was started in 2004 with the aim of generating tools for identifying all types of fish species. Its primary goal was to gather barcodes for all of the world's fish. FISH-BOL comprises sequences, geographical information, and images for examined specimens, thereby creating a valuable public resource. Information organized and analyzed through the BOLD database is later delivered via a data feed to the FISH-BOL web portal. This depository utilizes taxonomic information resulting from FishBase and maintains a catalogue of fish (Ward et al. 2008). The International Nucleotide Sequence Database Collaboration (INSDC) archives DNA sequences from the FISH-BOL campaign and annotates each sequence with the key word "barcode" when it meets the barcode data standards. It requires the bidirectionally sequenced 5'-end of the COI gene sequence, valid species name, details concerning voucher specimens, coordinates of the collection locality, collection date, collector, and identifier. Also required are a list of the primers used to generate reference sequences and archiving of the underlying electropherogram trace files in a publically accessible NCBI trace archive. All this information is useful for using barcodes in molecular diagnostics applications. BOLD provides an online workbench to FISH-BOL (Ward et al. 2008).

The FBIS web-based tool is designed for the fish of India. The database has a total of 2334 COI gene sequences belong to 472 aquatic species. It works both as a local DNA barcode library and as an analysis system and contains valuable data regarding the phenotype, distribution, and IUCN Red List status of fish (Nagpure et al. 2012). This database enables saving and extracting data in an easy way with simple steps. A user can submit species sequences through a submission protocol. Species identification is performed using similarity search programs; it finds homologues with almost 99% similarity to the query sequence, which accurately assigns the species (Nagpure et al. 2012).

14.3.3 NCBI GenBank

GenBank is a comprehensive database that contains nucleotide sequences for more than 250,000 species (Benson et al. 2013). NCBI offers an online/offline sequence submission platform to deposit sets of barcode sequences to the GenBank database. Along with the barcode data, the submission platform collects other annotations such as specimen voucher, geographical information, sample collection date, primer data, and raw files to help recognize the sequence's source organism and to maintain the accuracy of the sequence. The GenBank file structure format is easy to understand for users. It contains sequence data along with the accession numbers

and gene names, taxonomy, references to published literature, and other meaningful information. The GenBank format comprises the locus, definition, accession, keywords, source, reference, and features fields for the gene. The user can download the FASTA format nucleotide or amino acid sequence from the FASTA link given on files or send to menu option (<https://www.ncbi.nlm.nih.gov/genbank/barcode/>). It is important to give the publication details related to barcodes and sequences in FASTA format with reverse and forward primers. Protein sequence submission is optional.

14.4 DNA Barcoding Repositories and Their Associated Tools

It is difficult to preserve the data integrity, interoperability, and utility of information generated relating to the “what”, “where”, and “when” of biodiversity data. Furthermore, DNA barcoding and other biodiversity information systems must maintain data standards so that appropriate metadata is efficiently included. Three main organizations (the International Barcode of Life Project (iBOL), CBOL, and BOLD), promote barcoding research with the aim of generating reference barcodes (Group et al. 2009; Ratnasingham and Hebert 2007). These organisations are focused toward development of barcoding as a universal standard and offer an online workbench for collection, management, analysis, and use of DNA barcodes.

iBOL (<http://ibol.org>) has network of collaborators from about 150 countries, includes more than 190,000 marine species, and has identified 6000 potentially new species (flowering plants, ants, birds, butterflies, ants, mammals, bees, fish, and fungi). It has collections in the form of ecosystems such as rain forests, kelp forests, poles, seas, and coral reefs. CBOL generated the BOLD system as a catalogue of living beings and has collections covering more than 790,000 sequences, conforming to more than 67,000 correctly called “species.” The BOLD database entries contain barcode sequences and specimen information such as images, morphology, collection date, and geographical site. To provide practical utility for BOLD data, the mobile-based software DNA Barcoding Assistant efficiently maintains metadata for the gathering and management of specimen data for BOLD and other biodiversity information databases.

The DNA Barcoding Assistant (<http://www.dnabarcodingassistant.org/>) enables users to store and retrieve data such as provisional user-allocated taxonomic classification, geospatial data, digital images, and collection event information for specimens found in the field. Another web-based data-processing system tool, BioBarcode (<http://www.asianbarcode.org>), focuses on the collection of Asiatic organisms and encompasses about 11,300 specimen entries (Lim et al. 2009). On similar lines, a field information management system (FIMS) has been developed that provides information associated with tissues, collecting events, and specimens (Deck et al. 2012). Similarly, the Quick Response (QR) barcode system could be efficiently implemented to identify and track samples, together with relevant

information such as site details, time of collection, and taxonomic identity (Diazgranados and Funk 2013). These indicate that continuous progress is being made in DNA barcoding.

Acknowledgements Authors are thankful to the Director of CSIR-NEERI, Nagpur, Maharashtra, India for this work.

References

- Adrian-Kalchhauser I, Svensson O, Kutschera VE, Rosenblad MA, Pippel M, Winkler S, Schloissnig S, Blomberg A, Burkhardt-Holm P (2017) The mitochondrial genome sequences of the round goby and the sand goby reveal patterns of recent evolution in gobiid fish. *BMC Genomics* 18:177. <https://doi.org/10.1186/s12864-017-3550-8>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Baxevanis AD, Ouellette BF (2004) *Bioinformatics: a practical guide to the analysis of genes and proteins*, vol 43. Wiley, Hoboken
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. *Nucleic Acids Res* 41:36–42. <https://doi.org/10.1093/nar/gks1195>
- Cambiaghi A, Ferrario M, Masseroli M (2016) Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration. *Brief Bioinform* 12:bbw031. <https://doi.org/10.1093/bib/bbw031>
- Cawthorn DM, Steinman HA, Corli withuhn R (2012) DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Res Int* 46:30–40. <https://doi.org/10.1016/j.foodres.2011.11.011>
- Chen W, Ma X, Shen Y, Mao Y, He S (2015) The fish diversity in the upper reaches of the Salween River Nujiang River revealed by DNA barcoding. *Sci Rep* 5:17437. <https://doi.org/10.1038/srep17437>
- Costa FO, Landi M, Martins R, Costa MH, Costa ME, Carneiro M et al (2012) A ranking system for reference libraries of dna barcodes: application to marine fish species from Portugal. *PLoS One* 7:e35858. <https://doi.org/10.1371/journal.pone.0035858>
- Deck J, Gross J, Stones-Havas S, Davies N, Shapley R, Meyer C (2012) Field information management systems for DNA barcoding. *Methods Mol Biol* 858:255–267. https://doi.org/10.1007/978-1-61779-591-6_12
- Diazgranados M, Funk VA (2013) Utility of QR codes in biological collections. *PhytoKeys* 25:21–34. <https://doi.org/10.3897/phytokeys.25.5175>
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17:368–376
- Gao LM, Li Y, Phan LK, Yan LJ, Thomas P, Phan LK, Möller M, Li DZ (2017) DNA barcoding of east Asian Amentotaxus (Taxaceae): potential new species and implications for conservation. *J Syst Evol* 55:16–24. <https://doi.org/10.1111/jse.12207>
- Group CP, Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ (2009) A DNA barcode for land plants. *Proc Natl Acad Sci U S A* 106:12794–12797. <https://doi.org/10.1073/pnas.0905845106>
- Hanfling B, Lawson HL, Read DS, Hahn C, Li J, Nichols P, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Mol Ecol* 25:3101–3119. <https://doi.org/10.1111/mec.13660>
- Hebert PD, Gregory TR (2005) The promise of DNA barcoding for taxonomy. *Syst Biol* 54:852–859. <https://doi.org/10.1080/10635150500354886>
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proc Biol Sci* 270:313–332. <https://doi.org/10.1098/rspb.2002.2218>

- Henriques JM, da Costa Silva GJ, Ashikaga FY, Hanner R, Foresti F, Oliveira C (2015) Use of DNA barcode in the identification of fish species from Ribeira de Iguape Basin and coastal rivers from São Paulo state (Brazil). *DNA* 3:118–128. <https://doi.org/10.1515/dna-2015-0015>
- Higgs P, Manchester U (2001) Introduction to phylogenetics methods (ITP series on-line seminars). <http://online.kitp.ucsb.edu/online/infobio01/higgs/>
- Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, BurrIDGE M, Bernatchez L (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS One* 3:e2490. <https://doi.org/10.1371/journal.pone.0002490>
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755
- Ivanova NV, Zemplak TS, Hanner RH, Hebert PD (2007) Universal primer cocktails for fish DNA barcoding. *Mol Ecol Notes* 7:544–548. <https://doi.org/10.1111/j.1471-8286.2007.01748.x>
- Kuksa P, Pavlovic V (2007) Fast kernel methods for SVM sequence classifiers. In: Giancarlo R, Hannenhalli S (eds) *Algorithms in bioinformatics, Lecture Notes in Computer Science*, vol 4645. Springer, Berlin/Heidelberg, pp 228–239
- Kumar G, Kocour M (2017) Applications of next-generation sequencing in fisheries research: a review. *Fish Res* 186:11–22. <https://doi.org/10.1016/j.fishres.2016.07.021>
- Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9:299–306. <https://doi.org/10.1093/bib/bbn017>
- La Rosa M, Fiannaca A, Rizzo R, Urso A (2013) Alignment-free analysis of barcode sequences by means of compression-based methods. *BMC Bioinforma* 14:S4. <https://doi.org/10.1186/1471-2105-14-S7-S4>
- Landi M, Dimech M, Arculeo M, Biondo G, Martins R, Carneiro M, Carvalho GR, Brutto SL, Costa FO (2014) DNA barcoding for species assignment: the case of mediterranean marine fishes. *PLoS One* 9:e106135. <https://doi.org/10.1371/journal.pone.0106135>
- Lim J, Kim SY, Kim S, Eo HS, Kim CB, Paek WK, Bhak J (2009) BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources. *BMC Genomics* 10:1. <https://doi.org/10.1186/1471-2164-10-S3-S8>
- Little DP (2011) DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLoS One* 6:e20552. <https://doi.org/10.1371/journal.pone.0020552>
- Madden T (2013) The BLAST sequence analysis tool. In: *The NCBI handbook*. NCBI, Bethesda. https://unmc.edu/bsbc/docs/NCBI_blast.pdf
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? *PLoS Biol* 9:e1001127. <https://doi.org/10.1371/journal.pbio.1001127>
- Nagpure NS, Rashid I, Pathak AK, Singh M, Singh SP, Sarkar UK (2012) FBIS: a regional DNA barcode archival analysis system for Indian fishes. *Bioinformation* 8:483–488. <https://doi.org/10.6026/97320630008483>
- Pereira LH, Hanner R, Foresti F, Oliveira C (2012) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genet* 14:20–20. <https://doi.org/10.1186/1471-2156-14-20>
- Ratnasingham S, Hebert PD (2007) BOLD: the barcode of life data system (www.barcodinglife.org). *Mol Ecol Notes* 7:355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sbordoni V (2010) Strength and limitations of DNA barcode under the multidimensional species perspective. *Prog Probl*:271–276. ISBN 978-88-8303-295-0
- Shen Y, Guan L, Wang D, Gan X (2016) DNA barcoding and evaluation of genetic diversity in Cyprinidae fish in the midstream of the Yangtze River. *Ecol Evol* 6:2702. <https://doi.org/10.1002/ece3.2060>

- Shirak A, Dor L, Seroussi E, Ron M, Hulata G, Golani D (2016) DNA barcoding of fish species from the mediterranean coast of Israel. *Mediterr Mar Sci* 17:459–466. <https://doi.org/10.12681/mms.1384>
- Shokralla S, Hellberg RS, Handy SM, King I, Hajibabaei M (2015) A DNA mini-barcoding system for authentication of processed fish products. *Sci Rep* 5:15894. <https://doi.org/10.1038/srep15894>
- Steinke D, Vences M, Salzburger W, Meyer A (2005) TaxI: a software tool for DNA barcoding using distance methods. *Philos Trans R Soc Lond Ser B Biol Sci* 360:1975–1980. <https://doi.org/10.1098/rstb.2005.1729>
- Strauss R, Bond C (1990) Taxonomic methods: morphology. In: Schreck CB, Moyle PB (eds) *Methods for fish biology*. American Fisheries Society, Bethesda, pp 109–140
- Tanabe AS, Toju H (2013) Two new computational methods for universal DNA barcoding: a benchmark using barcode sequences of bacteria archaea animals fungi and land plants. *PLoS One* 8:e76910. <https://doi.org/10.1371/journal.pone.0076910>
- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, .. Gaboriaud C (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol Ecol* 25:929–942. <https://doi.org/10.1111/mec.13428>
- Velez-zuazo X, Alfaro-shigueto J, Mangel J, Papa R, Agnarsson I (2015) What barcode sequencing reveals about the shark fishery in Peru. *Fish Res* 161:34–41. <https://doi.org/10.1016/j.fishres.2014.06.005>
- Vera-Escalona I, Habit E, Ruzzante DE (2017) The complete mitochondrial genome of the freshwater fish *Galaxias Platei* and a comparison with other species of the genus *galaxias* (faraway, so close?). *Mitochondrial DNA* 28:176–177. <https://doi.org/10.3109/19401736.2015.1115497>
- Viswanathan R, Pillai VK (1956) Paper chromatography in fish taxonomy. *Proc Indian Acad Sci* 43:334–339. <https://doi.org/10.1007/BF03050245>
- Ward R, Hanner R, Hebert P (2008) The campaign to DNA barcode all fishes FISH-BOL. *J Fish Biol* 74:329–356. <https://doi.org/10.1111/j.1095-8649.2008.02080.x>
- Yang J, Vázquez L, Chen X, Li H, Zhang H, Liu Z, Zhao G (2017) Development of chloroplast and nuclear DNA markers for Chinese oaks (*Quercus* subgenus *Quercus*) and assessment of their utility as DNA barcodes. *Front Plant Sci* 8:816. <https://doi.org/10.3389/fpls.2017.00816>
- Yilmaz M, Yilmaz HR, Alas A (2007) An electrophoretic taxonomic study on serum proteins of *Acanthopoma Marmid* *Leuciscus Cephalus* and *Chondrostoma Regium*. *Eurasia J Biosci* 3:22–27
- Zhang J, Hanner R (2012) Molecular approach to the identification of fish in the South China Sea. *PLoS One* 7:e30621. <https://doi.org/10.1371/journal.pone.0030621>
- Zhang AB, Feng J, Ward RD, Wan P, Gao Q, Wu J, Zhao WZ (2012) A new method for species identification via protein-coding and non-coding DNA barcodes by combining machine learning with bioinformatic methods. *PLoS One* 7:e30986. <https://doi.org/10.1371/journal.pone.0030986>