eDNA Topic

# PCR: Finding or Designing Primers for your Species of Interest

Author: Stephen E. Harris, Ph.D.

## Introduction

Universal primers are designed to amplify a gene of interest across a wide variety of taxa. This approach, combined with advanced sequencing techniques, allows for the sequencing of all Polymerase Chain Reaction (PCR) products, enabling the detection of nearly all organisms within a sample. Beyond traditional DNA barcoding and biodiversity identification, this method can also reveal community interactions over large spatial scales. However, the design of primers for PCR is a critical factor in eDNA metabarcoding studies. Poorly designed primers can introduce biases in sequencing results due to variable coverage and resolution across taxa, which may lead to an inaccurate representation of biodiversity. Therefore, careful consideration of primer design is essential to ensure comprehensive and unbiased amplification of the target gene across the taxa of interest.

## Choosing the Right Gene Region for eDNA Studies

Several genes or gene regions (genetic loci) have become prominent in eDNA studies, primarily located within the mitochondrial and chloroplast genomes. These regions are widely used in DNA barcoding because they exist in multiple copies per cell and meet the DNA barcoding criteria of being identical within species yet variable between species. For instance, the mitochondrial gene **cytochrome oxidase I (COI)** has become the gold standard for metazoans, while the chloroplast gene **ribulose bisphosphate carboxylase large-chain (rbcL**) is commonly used for plants.

The choice of genetic loci can vary depending on the taxon and the available sequence data. However, it is essential to note that DNA barcoding is only effective when the sequence of interest can be compared against a well-curated reference database. Without such a database, a new one would need to be established to utilize the DNA barcode effectively.

In addition to COI and rbcL, other common DNA barcode regions are frequently used. For example:

- **12S** and **16S ribosomal RNA genes** are used for identifying fish and bacteria, respectively.

- **Internal transcribed spacer 1 and 2 genes (ITS1 and ITS2)** from nuclear ribosomal DNA are employed to identify fungi.

Barcode marker regions also have broader applications beyond the taxa typically associated with them. For example, due to the endosymbiotic origins of mitochondria, the 16S rRNA gene is also commonly used as a DNA barcode for vertebrates.

Once a DNA barcoding marker region is selected, the experimental designer must choose or design appropriate PCR primers to amplify the chosen marker region after DNA extraction. PCR is a fundamental technique in molecular biology, essential for amplifying specific DNA sequences from small amounts of DNA. This guide provides a concise introduction to the process of selecting or designing primers for eDNA studies, highlighting key decision points and practical considerations for ensuring successful amplification.

## Finding Primers for Species of Interest

One of the first steps in an eDNA study is to search for existing primers that have already been validated for the species of interest. This avoids the need to reinvent the wheel—if suitable, validated primers are available, you can skip the time-consuming and costly process of designing and testing new primers. Furthermore, using a well-established DNA barcoding locus increases the likelihood of finding an existing, well-populated reference database, which is crucial for minimizing bias. An underpopulated database may skew results by limiting taxonomic resolution.

Validated primers can often be found through:

- **Literature Search:** Conducting a thorough review of the scientific literature is key. Journals such as *Molecular Ecology Resources* and *Environmental DNA* frequently publish studies that include primer sequences. Using search engines like Google Scholar can help locate relevant publications.
- **Public Databases:** Online resources like PrimerBank, the Barcode of Life Data System (BOLD), and GenBank contain extensive information on validated primers and commonly used loci for various taxa. These databases allow users to search by keywords or species names to find primers that have been successfully applied in previous studies.
- **Collaborations:** After identifying relevant studies on DNA barcoding for the taxon of interest, reaching out to authors directly can be highly beneficial. Contacting them via email can provide insights into their work or lead to recommendations on DNA barcoding for particular taxa. This may also result in access to unpublished or proprietary primers that the authors are willing to share.

By leveraging these resources, you can streamline the primer selection process and focus on the analysis, increasing the efficiency and effectiveness of your eDNA study.

## Common Primers for Different Taxonomic Groups

### Animals (COI Gene):

#### Invertebrates

- Forward: LCO1490 and Reverse: HCO2198 (Folmer et al., 1994)
- Forward: dgLCO1490 and Reverse: dgHCO2198 (Meyer, 2003)

#### Vertebrates

- Forward: VF1d and Reverse: VR1d (Ivanova et al., 2007)
- Forward: Fish-BCL and Reverse: Fish-BCH (Baldwin et al. 2009)

#### Metazoans

- Forward: mICOIintF and Reverse: mICOIintR (Leray et al., 2013)

### Prokaryotes (16S rRNA gene)

- Forward: 27F and Reverse: 1492R (Weisburg et al., 1991)
- Forward: 515F (Parada et al., 2016) and Reverse: 806R (Apprill et al., 2015)

### Eukaryotes (18S rRNA gene)

- Forward: 1391F and Reverse: EukB (Stoeck et al., 2010)

### Fungi (ITS region)

- Forward: ITS1 and Reverse: ITS4 (Gardes and Bruns, 1993)

### Plants (various regions)

- Forward: *rbcl-a_F* and Reverse: *rbcl-a_R* (Kress et al., 2007)
- Forward: *matk_F* and Reverse: *matk_R* (Kress et al., 2007)
- Forward: *ITS2_F* and Reverse: *ITS2_R* (Kress et al., 2007)

| Primer Name | Direction | Sequence (5'->3') | Author | Uses |
|---|---|---|---|---|
| **LCO1490** | Forward | GGTCAACAAATCATAAAGATATTGG | Folmer et al., 1994 | Inverts, various |
| **HCO2198** | Reverse | TAAACTTCAGGGTGACCAAAAAATCA | Folmer et al., 1994 | Inverts, various |
| **dgLCO1490** | Forward | GGTCAACAAATCATAAAGAYATYGG | Meyer, 2003 | Inverts, various |
| **dgHCO2198** | Reverse | TAAACTTCAGGGTGACCAAARAAYCA | Meyer, 2003 | Inverts, various |

| | | | | |
|---|---|---|---|---|
| **VF1d** | Forward | TCTCAACCAACCACAARGAYATYGG | Ivanova et al., 2007 | Vertebrates |
| **VR1d** | Reverse | TAGACTTCTGGGTGGCCRAARAAYCA | Ivanova et al., 2007 | Vertebrates |
| **Fish-BCL** | Forward | TCAACYAATCAYAAAGATATYGGCAC | Baldwin et al. 2009 | Fish |
| **Fish-BCH** | Reverse | ACTTCYGGGTGRCCRAARAATCA | Baldwin et al. 2009 | Fish |
| **mlCOIintF** | Forward | TCAACYAATCAYAAAGATATYGGCAC | Leray et al., 2013 | Metazoan |
| **mlCOIintR** | Reverse | GGRGGRTASACSGTTCASCCSGTSCC | Leray et al., 2013 | Metazoan |
| **27F** | Forward | AGAGTTTGATCMTGGCTCAG | Weisburg et al., 1991 | Bacteria |
| **1492R** | Reverse | TACCTTGTTACGACTT | Weisburg et al., 1991 | Bacteria |
| **515F** | Forward | GTGYCAGCMGCCGCGGTAA | Parada et al., 2015 | Prokaryotes |
| **806R** | Reverse | GGACTACNVGGGTWTCTAAT | Apprill et al., 2015 | Prokaryotes |
| **1391F** | Forward | GTACACACCGCCCGTC | Stoeck et al., 2010 | Eukaryotes |
| **EukBr** | Reverse | TGATCCTTCTGCAGGTTCACCTAC | Stoeck et al., 2010 | Eukaryotes |
| **ITS1** | Forward | TCCGTAGGTGAACCTGCGG | Gardes and Bruns, 1993 | Fungi |
| **ITS4** | Reverse | TCCTCCGCTTATTGATATGC | Gardes and Bruns, 1993 | Fungi |
| ***rbcl-a_F*** | Forward | ATGTCACCACAAACAGAGACTAAAGC | Kress et al., 2007 | Plants |
| ***rbcl-a_R*** | Reverse | CTTCTGCTACAAATAAGAATCGATCTC | Kress et al., 2007 | Plants |
| ***matk_F*** | Forward | CCTATCCATCTGGAAATCTTAG | Kress et al., 2007 | Plants |
| ***matk_R*** | Reverse | GTTCTAGCACAAGAAAGTCG | Kress et al., 2007 | Plants |
| ***ITS2_F*** | Forward | TAGCTACTTCTTCGCAGC | Kress et al., 2007 | Plants |
| ***ITS2_R*** | Reverse | GGTCCAGTCCGCCCTGATGG | Kress et al., 2007 | Plants |

## Designing New Primers

If suitable primers are not available from existing sources, it becomes necessary to design new primers. This process involves several key steps and important considerations, both in silico (i.e., computationally) and in terms of the physical parameters affecting primer chemistry during PCR.

### Key Steps in Primer Design

#### 1. Target Sequence Selection

The first step is to identify a suitable target DNA region, or locus. Mitochondrial DNA and ribosomal RNA genes are commonly targeted. Both the DNA barcoding locus and its flanking regions should be carefully considered:

- **DNA Barcoding Locus:** The chosen genetic locus should be sufficiently variable between species but consistent within species. A locus that is too variable may hinder the effectiveness of DNA barcoding and could necessitate different primers across taxa.
- **Locus Flanking Regions:** The variable region used for species identification must have highly conserved flanking regions. These conserved regions enable universal primers to bind across a wide range of taxa.

Before designing primers, conduct a search of public databases to assess the number of available sequences for the target region in the taxa of interest. Download several reference sequences to use in primer design software.

#### 2. Primer Design Software

Several software tools are available to help design primers that meet important criteria, such as optimal melting temperature (Tm), GC content, and avoidance of secondary structures. Commonly used tools include:

- **Primer3:** A widely used tool for generating primer sets, allowing users to specify parameters such as target sequence, product size, and melting temperature.
- **OligoAnalyzer:** Analyzes primer properties, including melting temperature, GC content, and the potential for secondary structures like hairpins or dimers.
- **Primer-BLAST:** Combines primer design with a BLAST search to ensure primer specificity.
- **BioEdit:** A free sequence alignment editor and analysis tool designed for Windows.
- **Geneious:** A comprehensive bioinformatics platform that offers some free features, with upgraded capabilities available through a paid subscription.

### 3. *In Silico* Validation

Once a target region and primer sequences have been selected, it is essential to validate them computationally before synthesis. This step ensures specificity and efficiency. For example, Primer-BLAST can be used to align test primers against GenBank's large sequence database, confirming that the primers are compatible with target taxa and do not cross-react with non-target species.

Useful public-access reference databases for DNA barcoding include:

- **GenBank:** A comprehensive database of publicly available DNA sequences.
- **BOLD:** The Barcode of Life Data System, which serves as a repository for DNA barcodes.

By following these steps and carefully designing and validating primers, you can maximize the change of successful amplification of the target gene region for your eDNA study.

## Important Considerations in Primer Design

### 1. Specificity:

Primer design should align with the study's goals. Primers can be designed to target a single species with high specificity, reducing the risk of amplifying non-target DNA but potentially missing closely related species. Alternatively, universal primers can target a broader range of species, allowing for the detection of multiple taxa, though they may sacrifice some specificity.

### 2. Fragment Length:

The length of the PCR product (amplicon) is an important consideration. In eDNA studies, shorter amplicons are generally more efficient, particularly when working with degraded DNA. However, the amplicon length must still contain enough genetic variation to distinguish between species, balancing efficiency and taxonomic resolution.

## Physical and Chemical Parameters of Primer Design

This is an overview of chemical and physical characteristics that make a good primer, However, most primer design software tools already take these factors into consideration.

### Primer Length

**Optimal Length:** Primers should typically be 18–25 nucleotides in length. This range strikes a balance between specificity and efficiency. Shorter primers may be used, especially for eDNA studies, where DNA is often degraded and shorter sequences are easier to amplify.

### Melting Temperature (Tm)

Tm Range: Primers should have a melting temperature (Tm) between 55–65°C. Ideally, the Tm of the forward and reverse primers should be within 2–3°C of each other to ensure efficient annealing.

Calculation: The Tm depends on the nucleotide composition and the length of the primer. There are numerous online Tm calculators available to assist in determining the melting temperature.

### GC Content

Optimal GC Content: Primers should have a GC content (or guanine-cytosine content) of 40-60%, which ensures stable binding of the primer to the template DNA.

### Secondary Structures

Avoid Secondary Structures: Primers should be designed to avoid secondary structures such as hairpins, dimers, or other interactions that can interfere with PCR efficiency and primer binding.

### Product Size (Amplicon Length)

Optimal Length: In eDNA studies, the amplicon length should be relatively short (100–300 base pairs) to accommodate degraded DNA. However, for certain applications, lengths up to 600–800 base pairs can be acceptable.

### Degenerate Bases

Degenerate Primers: When designing primers for a group of related species, degenerate bases can be used to account for sequence variability. However, the number of degenerate bases should be minimized to prevent non-specific amplification. *In silico* testing can be used to optimize the use of degenerate primers.

## Primer Optimization for eDNA Metabarcoding

Whether a primer set is obtained from the literature or designed from scratch, it must be rigorously tested for compatibility with the specific protocols and goals of a study. Primer optimization is critical to ensure that PCR conditions are fine-tuned for the study's objectives.

In eDNA and metabarcoding studies, the primary goal of PCR is to selectively amplify the DNA of a particular species or group of species, enabling detection even when the target DNA is present in low abundance. A key objective of eDNA studies is to reduce false negatives, which occur when a species present in the sample is not amplified. Additionally, minimizing non-specific amplification—where DNA regions unrelated to the study's targets are amplified—is essential for ensuring accuracy.

To avoid such errors, primer optimization should begin early in the design process. Testing for ideal annealing temperatures, reaction conditions, and primer concentrations ensures that the primers effectively amplify the target species while avoiding the amplification of non-target sequences. This step is crucial for maximizing the sensitivity and specificity of the assay and ensuring that the study accurately detects the presence of species in environmental samples.

## Considerations During Primer Optimization

Note that no matter how well-designed a primer set may be, it can still introduce bias in results due to non-uniform amplification or non-specific binding and amplification. Using multiple primer sets and cross-checking results may help mitigate this issue. However, increasing the number of primer sets, samples, and PCR replicates can significantly raise costs.

To reduce costs when testing primers for next-generation sequencing (NGS), primers can be tagged with unique nucleotide sequences (i.e., indexes) to identify sequencing reads that originate from specific samples, replicates, or primer sets. This process, known as multiplexing, helps lower the cost of testing primers for NGS. Moreover, once the primers are finalized, multiplexing can also reduce the overall cost of sequencing large numbers of samples.

It is also important to note that although flanking regions of a DNA barcoding locus should be highly conserved, genetic variation in these regions can exist across widely divergent taxa. This variation may prevent universal primer binding. To make primers more universal, researchers can incorporate degeneracy into their design. Traditionally, primer sets would include multiple oligonucleotides with different nucleotides at specific positions, increasing the likelihood of primer binding across a broader range of taxa. Today, synthetic nucleotides are available for primer synthesis that can bind to multiple nucleotides. These synthetic nucleotides are represented by letters other than A, T, C, or G in primer sequences. Below is a summary of these nucleotide codes, adapted from Srivathsan et al. (2021):

| **Degenerate nucleotide code** | |
|---|---|
| **W = A or T** | B = C or G or T |
| **S = G or C** | D = A or G or T |
| **M = A or C** | H = A or C or T |
| **K = G or T** | V = A or C or G |
| **R = A or G** | N = A or C or G or T |

## Example Protocol 1:
## Searching the Literature and Finding a Relevant Primer for Your Taxon of Choice

### Step 1: Define your target taxon

1. Identify the taxon of interest (e.g., species, genus, family).
2. Clearly define the specific objective of your study (e.g., detection of a non-native fish species in a lake).

### Step 2: Conduct a literature search

1. Use search engines like Google Scholar, PubMed, and Web of Science.
    - Use keywords such as "eDNA primers for [taxon name],", "PCR primers for [species name],", or "DNA barcoding [genus]". This will help pinpoint what genetic locus is most commonly used for your species and help you find primers that may already be available.
2. If you find review articles or meta-analyses, these can help in summarizing the field and identifying some efficient primers that are currently in use.

### Step 3. Access online databases

Once you have a primer of interest, you should test it *in silico* against a reference sequence. The following resources are good for testing specific primers against your species of interest.

- **GenBank:** Search the GenBank nucleotide database for sequences related to your taxon.
    - Use the Nucleotide database with terms like "[taxon name] COI", or "[species name] 16S.".
- **Barcode of Life Data System (BOLD):** Search for barcodes and primers specific to your taxon.
    - Navigate to the BOLD Systems website and use the search feature.
- **PrimerBank:** Look for validated primers in PrimerBank.
    - Enter your taxon or specific genes (e.g., COI, 16S) in the search bar.

### Step 4: Extract primer information

Record primer sequences, references, and target regions. Make sure to include the directionality, (i.e., the 5’ and 3’ direction).

Example format:

- **Forward primer:** LCO1490, sequence: 5'-GGTCAACAAATCATAAAGATATTGG-3'

- **Reverse primer:** HCO2198, sequence: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3'

### Step 5: Evaluate primer performance

For this step, you will want to test your primer against as many sequences of your species of interest as possible.

1. *In Silico* testing of a specific primer:
    - If you are attempting to identify one specific species, you should download as many sequences of that species as possible for your given gene, as well as several sequences from sister species from NCBI GenBank.
    - Use Primer-BLAST to evaluate the specificity of the primers against these reference sequences.
    - Note that Primer-BLAST does not handle degeneracy, but you can change each nucleotide to the 'N' character as a substitute.

2. *In Silico* testing of a primer set against broad multiple sequences:

    For this step, you will want to test your primer against as many sequences of your species of interest as possible.

    - If you are interested in identifying a certain family (e.g., Ginglymostomatidae, the Nurse Sharks), you should download a handful of sequences from each species within your group of interest as well as several outliers from NCBI GenBank. You will be testing your primers against these sequences for compatibility and to ensure there is not unacceptable, non-specific binding. This is good for checking universal primers against a large group.
    - Input all your sequences in BioEdit or Geneious
    - Perform a ClustalW multisequence alignment
    - Input your primer DNA sequences as forward and reverse FASTA sequences
    - Use the pPrimer tool in Geneious to 'Test with Saved Primers' and ensure your primers align to the sequences of interest without too much non-specific binding.

## Example Protocol 2:
## Designing Your Own Primer Using Software

### Step 1: Collect Sequences:

1. **Retrieve Sequences:** Gather DNA sequences for your target species from databases like GenBank or BOLD.
2. **Align Sequences:** Use software like MEGA or Clustal Omega to align sequences and identify conserved regions.
    - Save the alignment in a format compatible with primer design software (e.g., FASTA).

### Step 2: Design Primers Using Primer3

1. **Access Primer3:** Open the Primer3 web interface or install Primer3 on your local machine.
2. **Input Sequence:** Paste the aligned sequence into the Primer3 input box.
3. **Set Parameters:**
    - **Target region:** Define the start and end of the target region.
    - **Product size:** Set the desired amplicon length (e.g., 100–300 bp).
    - **Primer properties:** Adjust settings for optimal Tm (55–65°C), GC content (40–60%), and absence of secondary structures.

### Step 3: Analyze Primer Candidates

1. **Primer Evaluation:** Examine the suggested primers for specificity and efficiency.
2. ***In Silico* Validation:** Use Primer-BLAST to ensure primers specifically amplify the target region without cross-reacting.

## Frequently Asked Questions

1. What are the essential criteria for selecting a good primer?
   A good primer should have the following properties:

    - **Optimal length:** 18–25 nucleotides (shorter is okay for eDNA primers).
    - **Melting temperature (Tm):** Ideally between 55–65°C.
    - **GC content:** 40–60% is generally preferred.
    - **Minimal secondary structures:** Avoid hairpins and self-complementarity in general.
    - **Specificity:** Primers should specifically bind to the target sequence without significant non-specific binding to other regions or species/taxon not of interest.

2. How can I verify the specificity of my designed primers?

*In silico* verification using tools like Primer-BLAST, BioEedit, and Geneious can help ensure specificity. Primer-BLAST compares your primers against a nucleotide database to check for potential cross-reactivity with non-target sequences.

For a detailed protocol, see Ye et al. (2012), which provides a comprehensive guide on using Primer-BLAST.

3. What is the difference between degenerate and non-degenerate primers?

Degenerate primers contain multiple possible nucleotides at certain positions, allowing them to bind to a variety of related sequences. Non-degenerate primers have a fixed nucleotide sequence. Degenerate primers are useful when amplifying DNA from a group of related organisms, while non-degenerate primers are more specific to a single sequence.

4. How can I optimize PCR conditions for my primers?

PCR optimization with primer design and relies on annealing temperature (which is calculated from primer melting temperature), primer concentration, primer length, and degeneracy.

5. What is the best way to make sure my primers are going to work?

In order to avoid wasting time and money, primers should be tested *in silico* before synthesizing. *In silico* testing helps identify potential issues with primer specificity and efficiency. Refer to “Example Protocol 2” for *in-silico* testing guidance and see the Primer3 software user manual and the Primer-BLAST documentation.

6. What are some common issues with eDNA primers, and how can they be resolved?

Common issues include non-specific amplification, primer-dimer formation, poor amplification efficiency, and bias in taxon amplification. These can often be resolved by redesigning the primers, optimizing PCR conditions, or using additives that help deal with environmental inhibitors, such as bis(trimethylsilyl)acetamide (BSA).

7. How do I select the target region for primer design in eDNA studies?

The target region should be conserved across the species of interest, but variable enough to distinguish between different species. Common targets include mitochondrial COI for animals, 16S rRNA for bacteria, and ITS regions for fungi.

8. Can I use universal primers for eDNA analysis?

Yes, universal primers can amplify a broad range of species within a taxonomic group, making them useful for biodiversity studies. However, they may have lower specificity compared to species-specific primers.

9. How can I validate my primers experimentally?

Experimental validation involves testing the primers on DNA from samples that are known to contain target DNA (i.e., positive control) and at least one sample that does not per PCR (i.e., negative control). Sensitivity and specificity should be evaluated by testing the primers on a range of environmental samples.

## Further reading

Apprill A, McNally S, Parsons R, Weber L (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquatic Microbial Ecology, 75, 129-137. https://doi.org/10.3354/ame01753

Baldwin CC, Mounts JH, Smith DG, Weigt LA (2009) Genetic identification and color descriptions of early life-history stages of Belizean Phaeoptyx and Astrapogon (Teleostei: Apogonidae) with comments on identification of adult Phaeoptyx. Zootaxa, 22, 1–22. https://www.mapress.com/zt/article/view/zootaxa.2008.1.1

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994 Oct;3(5):294-9. https://www.mbari.org/wp-content/uploads/2016/01/Folmer_94MMBB.pdf

Gardes M, Bruns TD. ITS primers with enhanced specificity for basidiomycetes--application to the identification of mycorrhizae and rusts. Mol Ecol. 1993 Apr;2(2):113-8. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x.

Ivanova N V., Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. Molecular Ecology Notes, 7, 544–548. https://doi.org/10.1111/j.1471-8286.2007.01748.x

Kress WJ, Erickson DL. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One. 2007 Jun 6;2(6):e508. https://doi.org/10.1371/journal.pone.00005088.

Leray, M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for

metabarcoding metazoan diversity: application for characterizing coral. https://doi.org/10.1186/1742-9994-10-34

Meyer, C. P. (2003). Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. Biological Journal of the Linnean Society, 79(3), 401-459. https://doi.org/10.1046/j.1095-8312.2003.00197.x

Parada AE, Needham DM, Fuhrman JA (2015) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology, 18, 1403-1414. https://doi.org/10.1111/1462-2920.13023