

## eDNA Topic

# Sequencing: Technologies and Multiplexing

Author: Jonah Ventures

## Introduction

DNA sequencing refers to a range of molecular methods used to determine the specific order of nitrogenous bases (A, T, C, and G) in DNA. While traditional “first-generation” sequencing, such as Sanger sequencing, has been widely used, environmental DNA (eDNA) studies typically rely on high-throughput or Next Generation Sequencing (NGS) methodologies. NGS includes both “second-generation” technologies for sequencing short DNA fragments (e.g., Illumina) and “third-generation” technologies for sequencing longer fragments (e.g., PacBio and Oxford Nanopore). These methods are “high-throughput,” meaning they allow large volumes of DNA to be sequenced simultaneously, ranging from genomic libraries of environmental samples to entire genomes.



## The Role of Sequencing in Metabarcoding

Metabarcoding is a powerful tool in eDNA studies, allowing scientists to identify and quantify species in an ecosystem using environmental samples like water, air, or soil. DNA sequencing plays a central role in this process. To detect rare DNA sequences, PCR (Polymerase Chain Reaction) is used to amplify these DNA fragments. In quantitative PCR (qPCR), the abundance of a single sequence can be measured. Metabarcoding, however, extends this capability by enabling the simultaneous sequencing of multiple DNA fragments from different species within a single sample. These fragments, called DNA barcodes, represent unique patterns of DNA sequences specific to particular species. The amplified sequences are then processed by high-throughput sequencing platforms, generating data files for subsequent bioinformatic analysis.

## Sequencing: Long- vs. Short-read Technologies

Sequencing technologies can be categorized based on the length of the sequence reads they produce. Both short- and long-read sequencing methods have value for eDNA studies, depending on the study and the application of results. In order to decide which tool is appropriate for a study, an experimental designer must understand how each is performed and applied.

## Short-read Sequencing

Short-read sequencing, most commonly performed on Illumina platforms, is known for its high accuracy and throughput.

### Illumina Sequencing

Illumina sequencing works by sequencing by synthesis, where fluorescently labeled nucleotides are incorporated into a growing DNA strand, and each incorporation is detected in real-time. This technology can generate read lengths of up to 600 base pairs (bp) over roughly two days. For Illumina sequencing, DNA fragments (amplicons) must be prepared with adapters through PCR. These libraries include DNA inserts flanked by adapter sequences, typically adding about 120-150 bp. The adapters contain P5 and P7 sequences for flow cell binding, unique indexes for sample identification, and binding sites for sequencing primers. Both forward (Read 1) and reverse (Read 2) reads are generated and merged during bioinformatic processing. The longest amplicons sequenced by Illumina (excluding adapters, indexes, and overlaps) are typically around 400 bp.

### Long-read Sequencing

Long-read sequencing technologies allow the sequencing of DNA fragments ranging from thousands of base pairs to entire chromosomal lengths. Two prominent long-read sequencing technologies are offered by Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT).

### PacBio Sequencing

PacBio's Single Molecule, Real-Time (SMRT) sequencing can generate reads spanning thousands of base pairs, making it ideal for complex genome regions and *de novo* assembly. In metabarcoding, PacBio allows for sequencing longer barcodes, improving species identification accuracy. PacBio's sequencing works by detecting fluorescent signals emitted during nucleotide incorporation as DNA polymerase synthesizes a complementary strand to the template DNA. This real-time detection allows for long DNA fragments to be sequenced, reducing the need for assembly.

### Nanopore Sequencing

Nanopore sequencing from Oxford Nanopore Technologies (ONT) involves passing DNA strands through nanopores embedded in a membrane. ONT flow cells contain hundreds to thousands of nanopores, each capable of sequencing a single DNA molecule. As DNA moves through the nanopore, it disrupts ion flow, causing changes in voltage that are characteristic of the nucleotide sequence. These changes are detected in real-time and converted into a base

sequence. Nanopore sequencing does not require DNA synthesis and can read long DNA fragments, enabling sequencing of long amplicons. A motor protein attached to the DNA strand regulates its movement through the nanopore at a manageable speed, powered by ATP. Nanopore sequencing can read up to 400 bases per second and can deliver complete amplicon data in less than ten minutes.

### Multiplexing Strategies in Sequencing

Running a single sample at a time with either short- or long-read sequencing technologies would be cost-prohibitive. To address this, multiplexing strategies are used to allow simultaneous sequencing of multiple samples in a single run. Multiplexing involves indexing, where unique sequences (indexes) are added to each sample. Indexing can be done by adding a unique index to just one read (typically the forward read) or by using dual indexing, where both the forward and reverse reads are indexed. Unique dual indexing uses distinct index pairs for each sample, ensuring that no individual index is reused. This method provides higher accuracy and significantly reduces the risk of misassigning sequences to the wrong sample, which is crucial for maintaining the integrity of results in large-scale sequencing projects.

### Pooling and Normalization of Amplicons

Once the indexing PCR is complete, amplicons from different samples must be pooled for sequencing. Before pooling, normalization of DNA concentrations ensures that each sample contributes equally to the sequencing pool by adjusting the concentration of each amplicon to a standard level. Common techniques for normalization include:

- **Bead-based Normalization:** This method uses magnetic beads to bind and equalize DNA concentrations across samples, ensuring uniform input for sequencing.
- **Spectrophotometry:** This technique measures the absorbance of DNA at specific wavelengths to accurately quantify DNA concentrations, allowing them to be adjusted as needed.

During pooling, it is crucial to account for the primers and indexes used in different samples to ensure they can be correctly separated during data analysis. Amplicons with the same indexes can be mixed if generated with different primers. However, if amplicons from different samples share the same primers and indexes, they cannot be bioinformatically separated after sequencing, which could compromise the integrity of the data. Therefore, careful planning is essential to avoid such conflicts during pooling.

---

## Example Protocol: Steps for Indexing (Short-read) Amplicons

---

### 1. Prepare PCR mix for indexing:

Use Illumina's Library Prep and Index adapter kits to prepare a PCR mix for each sample that includes the amplicon DNA, unique dual index (UDI) primers (both forward and reverse), and polymerase master mix.

### 2. Run PCR to attach UDI primers

Follow the protocol-specific instructions run the PCR with UDI primers.

### 3. Clean-up and normalize PCR products

- **Bead-Based Normalization:** This method uses magnetic beads to bind and equalize DNA concentrations across samples, ensuring uniform input for sequencing.
- Add an equal volume of AMPure XP magnetic beads to each sample.
- Incubate the samples at room temperature for 5 minutes.
- Place the samples on a magnetic stand and wait until the solution clears.
- Carefully remove and discard the supernatant, ensuring not to disturb the pelleted beads.
- Wash the beads with 200  $\mu$ L of 70% ethanol. Repeat the wash step.
- Air dry the beads for 5–10 minutes.
- Elute the DNA with a suitable volume of elution buffer to normalize the concentration (e.g., 20  $\mu$ L).

### 4. Pool indexed amplicons by combining equal volumes of sample into a single tube.

### 5. Quantify the final pooled library (collection of DNA sequences and adapters) using a fluorometric method, such as a Qubit or NanoDrop.

### 6. Check the size distribution using an electrophoresis system (e.g., Bioanalyzer or TapeStation) in order to calculate final loading volumes

### 7. The final pooled library is now ready for loading onto the Illumina sequencer according to the manufacturer's instructions.

## Frequently Asked Questions

---

### 1. Is it sufficient to have just one index for an amplicon, or are dual indexes necessary?

Using a single index can be sufficient in many cases, but dual indexes are generally recommended for greater accuracy and reduced cross-contamination. Dual indexes, placed on both ends of the amplicon, provide a unique identifier for each sample, significantly minimizing the risk of misidentification, especially in high-throughput sequencing projects where multiple samples are multiplexed. While purchasing dual indexes is more expensive, the added cost is often justified in labs handling large datasets or where high accuracy is critical. However, for smaller projects or labs with limited resources, single indexing may still be a practical choice.

### 2. If I use dual indexes, do both the forward and reverse indexes need to be unique, or can just one be unique?

Using unique dual-indexing, where both the forward and reverse indexes are unique for each sample, is the most effective way to ensure accurate sample identification. This method greatly reduces the likelihood of misassignment and improves the reliability of sequencing results. When multiplexing large numbers of samples, combining forward and reverse indexes in a factorial manner is a cost-efficient way to achieve a high level of multiplexing. For instance, using 96 forward indexes and 96 reverse indexes can generate 9216 unique dual-index (UDI) combinations. However, generating over 9000 UDIs can be expensive, potentially costing upwards of \$100k, making this an investment that may only be justified in high-throughput labs or large-scale projects.

### 3. Are short reads sufficient for my eDNA study, or should I use longer reads?

Short reads are typically sufficient for eDNA metabarcoding studies, especially when targeting specific regions like the 12S or 18S rRNA genes, which are variable enough to allow good taxonomic discrimination using short-read platforms such as Illumina. However, longer reads can offer better taxonomic resolution, particularly for closely related species, and may reduce the need to use multiple markers. That said, some eDNA samples may not contain enough long DNA sequences to reliably amplify, limiting the practicality of long-read sequencing in certain cases.

### 4. What are the drawbacks of using long-read sequencing platforms over Illumina?

Long-read sequencing platforms like Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) offer the advantage of reading longer sequences, which can be particularly useful for more comprehensive genomic studies and improved taxonomic resolution. However, long-read platforms have traditionally been associated with higher

error rates compared to Illumina, which can affect the accuracy of species identification. That said, the gap in accuracy is closing rapidly as technologies improve. While long-read sequencing is generally more expensive per base of sequence data, this cost disparity is also narrowing. As a result, the field of eDNA metabarcoding may transition to using longer reads in the near future, but the timeline and adoption of this shift remain uncertain.

#### 5. I heard that Nanopore sequencing can be done in the field. Is this a good idea?

Nanopore sequencing using the portable MinION device can be conducted in the field, offering the advantage of real-time data generation and immediate analysis. However, field conditions can pose challenges, such as equipment stability, environmental impacts, and less controlled sample preparation, which may increase the risk of contamination. While nanopore sequencing enables rapid results, it generally has a higher error rate compared to lab-based Illumina sequencing. Additionally, performing base calling (converting electrical signals into DNA bases) in the field can be difficult, potentially affecting the accuracy of species detection and identification. Balancing these trade-offs is crucial, as field-based sequencing provides flexibility but may come at the cost of precision and control.

#### 6. What is a spike-in and is it necessary for sequencing eDNA samples?

A spike-in is a sequencing control that adds a diversity of bases to the sequencing library. Usually, the PhiX bacteriophage genome is used, but other spike-ins are necessary for different applications. PhiX is mainly needed when the library pool has low diversity, where most of the DNA sequences are the same. For those low diversity libraries, PhiX can be spiked-in to the library to make up to 40% of the library's total concentration. For high-diversity libraries, a 1% spike-in of PhiX is sufficient.,

### Further reading

---

Bohmann, K., Elbrecht, V., Carøe, C., Bista, I., Leese, F., Bunce, M., Yu, D.W., Seymour, M., Dumbrell, A.J. and Creer, S., 2022. Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources*, 22(4), pp.1231-1246.

<https://doi.org/10.1111/1755-0998.13512>

Rishan, S.T., Kline, R.J. and Rahman, M.S., 2023. Applications of environmental DNA (eDNA) to detect subterranean and aquatic invasive species: a critical review on the challenges and limitations of eDNA metabarcoding. *Environmental Advances*, p.100370.

<https://doi.org/10.1016/j.envadv.2023.100370>

Takahashi, M., Saccò, M., Kestel, J.H., Nester, G., Campbell, M.A., Van Der Heyde, M., Heydenrych, M.J., Juskiewicz, D.J., Nevill, P., Dawkins, K.L. and Bessey, C., 2023. Aquatic environmental DNA: A review of the macro-organismal biomonitoring revolution. *Science of the Total Environment*, 873, p.162322. <https://doi.org/10.1016/j.scitotenv.2023.162322>

van der Reis, A.L., Beckley, L.E., Olivar, M.P. and Jeffs, A.G., 2023. Nanopore short-read sequencing: A quick, cost-effective and accurate method for DNA metabarcoding. *Environmental DNA*, 5(2), pp.282-296. <https://doi.org/10.1002/edn3.374>