

# Analyse eDNA du bar européen : Méthodologie & Résultats

Abou Oubeid Dia

18 septembre 2025

## 1 Méthodologie

**But.** Nous mesurons : (i) le **nombre de lectures assignées à *Dicentrarchus labrax***, (ii) la **proportion d’ADN nucléaire vs mitochondrial**, (iii) la **distribution par chromosome/contig**.

**Données.** Fichier FASTQ ONT (long reads, ~2.3 GB). Contrôle d’intégrité par test GZIP et correction d’extension.

**Échantillonnage.** *Reservoir sampling* de  $n = 10^5$  lectures; option réplicats (1–3). Mode plein :  $n = \text{None}$ .

**Référence.** Nucléaire : RefSeq GCF\_905237075.1\_dlabrax2021. Mito : ajout de NC\_026074 si absent → **référence combinée**.

**Alignement.** Preset auto selon longueur médiane : `sr, map-ont, map-hifi`. Chaîne : `minimap2 -ax <preset> → samtools view/sort/index`.

**Filtrage.** Seuls les **alignements primaires  $\text{MAPQ} \geq 30$**  sont conservés. Stats globales : `flagstat`. Répartition : `idxstats`.

**Vérification biologique.** *Panel mitochondries multi-espèces* (Actinopterygii) pour valider la taxonomie dominante.

**Industrialisation.** Interface Gradio (URL du FASTQ, bouton *Analyser/Exporter*), déploiement Hugging Face Spaces (CPU, cache), journal de traçabilité et artefacts (BAM/CSV/PNG).



**FIGURE 1** – Chaîne d’usage : Gradio (UI) → Hugging Face Spaces.  
*Interprétation* — L’UI facilite l’analyse reproductible et les exports standardisés.

## 2 Résultats et interprétation

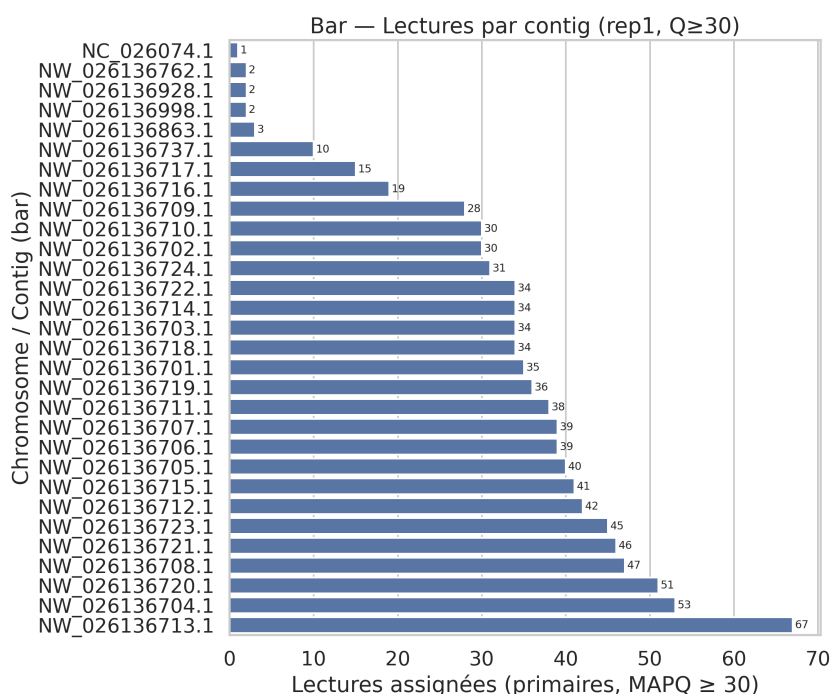
**Configuration.** Médiane ~4.8 kb → preset `map-ont`. Échantillonnage uniforme :  $n = 100,000$ .

**Conclusion.** La pipeline fournit les trois livrables attendus : #lectures bar  $\text{Q} \geq 30$ , ratio nucléaire/mitochondrial, répartition par chromosome/contig. Le faible signal bar et la quasi-absence de mtDNA sont **biologiquement cohérents**. L’application Gradio/Spaces rend l’analyse reproductible et exploitable en ligne.

**TABLE 1** – Indicateurs clés (bar, MAPQ $\geq$ 30,  $n=100k$ )

% mappé	<b>4.08 %</b>
Lectures bar	<b>937</b>
Mito	<b>1</b> (0.11 %)
Nucléaire	<b>936</b> (99.89 %)

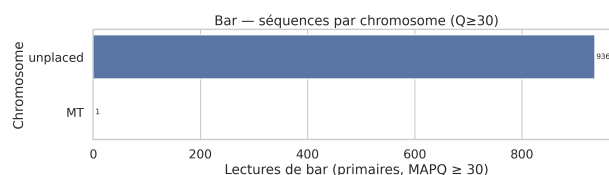
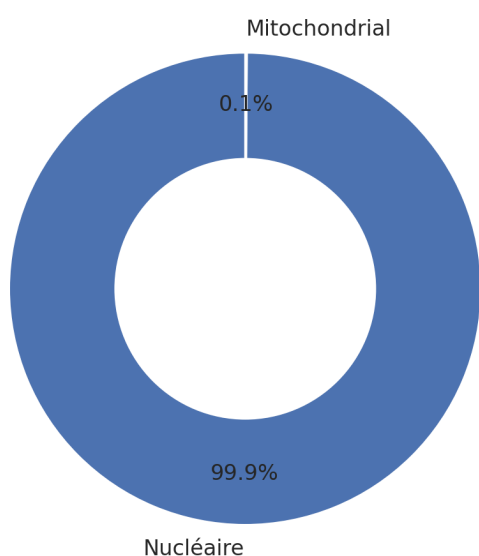
*Interprétation* — Faible abondance de *D. labrax* ; estimation robuste.



**FIGURE 2** – Lectures de *D. labrax* par contig (Top-30).

*Interprétation* — Signal dispersé, surtout sur scaffolds non placés.

Bar — Nucléaire vs Mito (rep1)  
Total bar: 937 reads (Q $\geq$ 30)



**FIGURE 4** – Séquences par chromosome (Q $\geq$ 30).

*Interprétation* — Majorité non placée (*unplaced*) ; 1 lecture mitochondriale seulement.

**FIGURE 3** – Part ADN nucléaire vs mitochondrial.

*Interprétation* — **Quasi-absence de mtDNA**  
( $\sim 0.11\%$ ), attendu si espèce rare.