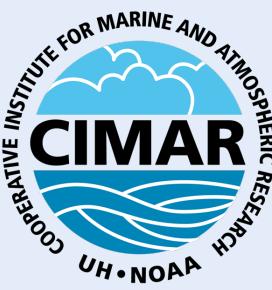


Reference Databases for eDNA & Metabarcoding analyses

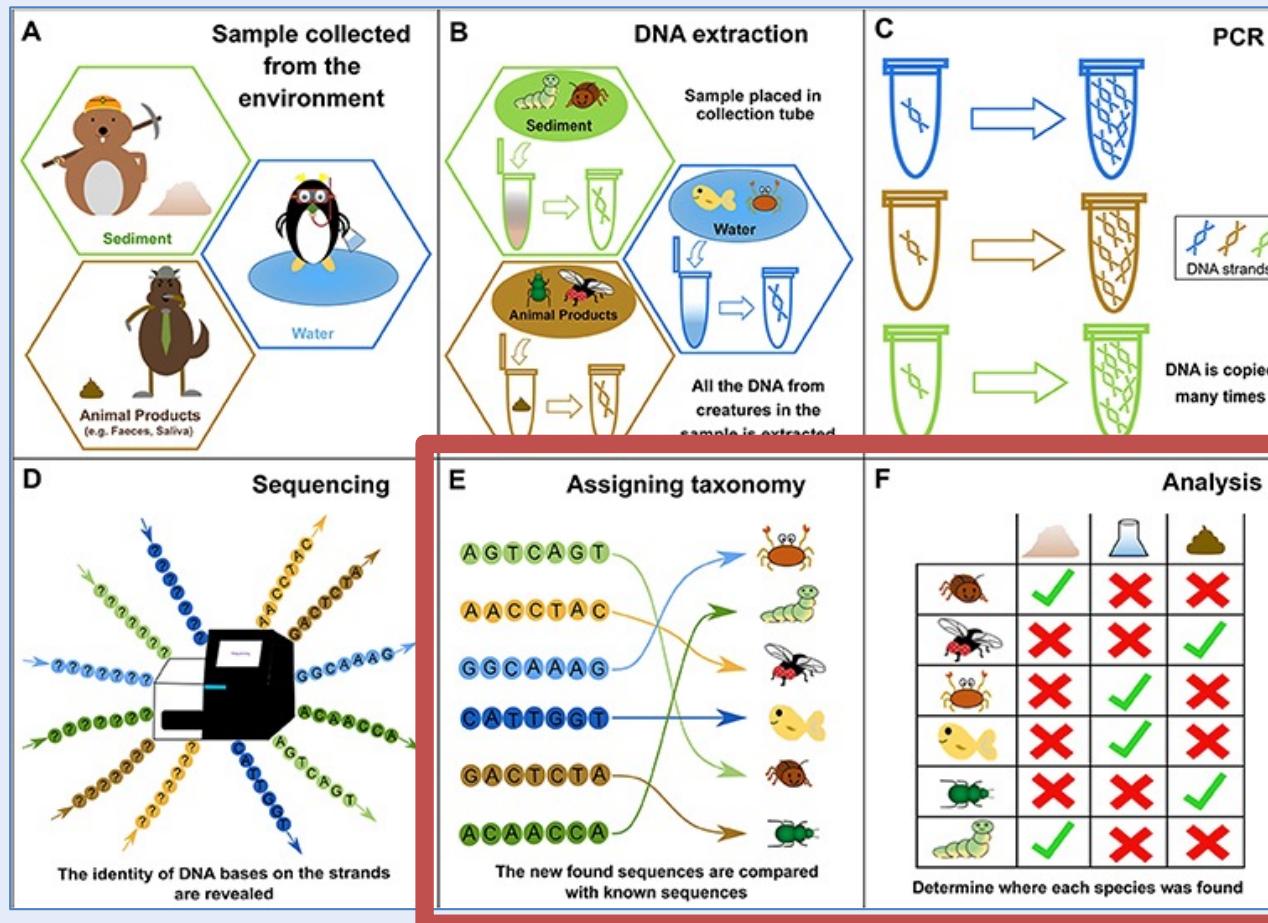
Concepts and Resources Available in SEDNA

Dr. Eric Garcia

eric.e.garcia@noaa.gov



Metabarcoding Study



Schallenberg et. al. 2020 doi: 10.3389/frym.2019.00150

Reference quality heavily impacts taxonomic assignment and downstream analyses

Why Reference Databases Are Critical

- They define what taxa can be detected
- Errors propagate into ecological conclusions
- Incomplete databases cause false absences
- Curation level directly impacts confidence

Database and Assignment Caveats

- **Assignment Errors:** misannotation e.g. incorrect species or genus assignments
- **Gaps:** poorly studied, cryptic or rare taxa might not be fully represented or absent
- **Biases:** model taxa, economically important, or of high interest have better representation
 - Databases created for specific groups

Common Reference Databases

SEDNA

NCBI

GenBank (nt)

NCBI GenBank / nt: comprehensive but minimally curated

BOLD (COI)

BOLD: barcode-focused, strong for COI but taxonomically uneven

**SILVA
(16S/18S)**

SILVA: ribosomal RNA (16S/18S), microbial and eukaryotic focus

PR2 / UNITE

PR2 & UNITE: protists and fungi, respectively

SEDNA

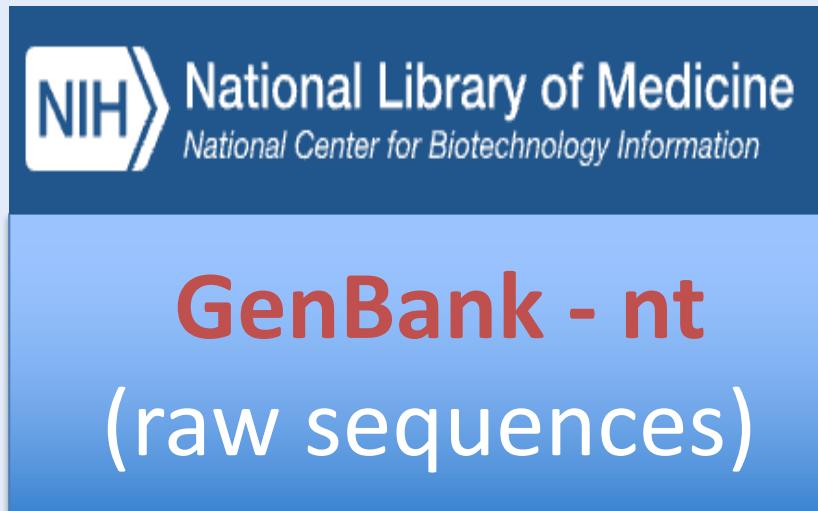
**MIDORI2
(mtDNA)**

MIDORI2: curated mitochondrial database for eukaryotes

Limitations of Using NCBI nt Directly

- Contains misidentified or outdated taxonomic annotations
- Redundant sequences and inconsistent taxonomic ranks
- Includes environmental, low-quality, and partial records
- Not optimized for metabarcoding workflows

MIDORI2: A Curated Alternative



- Curated
- Filtered
- Standardized
- Versioned

MIDORI2 Scope and Markers

- Mitochondrial genes: COI, 12S, 16S, Cytb, ND genes, ATP6/8
- 13 protein-coding + 2 rRNA markers
- Covers Metazoa, plants, fungi, algae, and protists
- Designed for broad-range eDNA primers and metabarcoding

How MIDORI2 Handles Taxonomy

NCBI Taxonomy IDs included

Inaccurate – mislabeled
assignments removed

Missing ranks inferred from
lower levels

Why MIDORI2 Is Valuable (Evidence)

- Broader taxonomic coverage than other databases (i.e. BOLD and CO-ARBitrator)
- Better performance for mitochondrial markers
- BLAST+ top-hit performs well with MIDORI2
- Regular updates and versions ensure long-term relevance

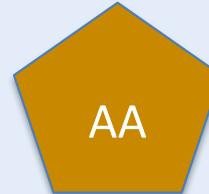
When MIDORI2 Is the Right Choice

- COI, 12S, 16S metabarcoding studies
- Broad eukaryotic community surveys
- Long-term monitoring requiring version control
- Reproducible, publication-ready workflows

MIDORI2 Database Options



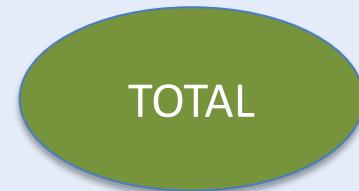
any format



or



Include "sp.," "aff.," "nr.," "cf.," "complex," etc
YES or **NO**





SEDNA – NCBI – MIDORI2

This README documents general information and provides a loose guide to help you work on NOAA's supercomputer **SEDNA**.

Please read the [SEDNA information and best practices](#)

Many of

What

You're pl
or storag
that live

Working on SEDNA

Sequence Databases

There are a few sequence databases (NCBI and MIDORI2) that are available to all SEDNA users.

NCBI nt Database

I have installed the lastest NCBI nt database at:

```
/share/all/ncbi_databases/
```

where the date is the day I installed it.

Here some info:

Midori2 SEDNA Databases

This README documents general information about Midori2 Databases and provides a loose guide to help you use, setup or update these on NOAA's supercomputer **SEDNA**

Note: if you are new to SEDNA or still need general help to work in SEDNA, please start by reading the [SEDNA information and best practices](#) and/or the [Working on SEDNA README](#)

NCBI GenBank databases are known to have various problems such as erroneous identification of organisms, potential lack of sequence curation, etc. This is where Midori2 can be useful for metabarcoding analyses.

Midori2 is a set of publicly accessible, mitochondrial marker or amino acid databases (from NCBI GenBank), that have already been curated and gets updated every few months. In addition, these databases have also been pre-formatted to fit many common metabarcoding pipelines, and raw sequences are also available if your desired format is not included.

Create a new release

Packages

No packages published

[Publish your first package](#)

Contributors 2

 ericgarciaresearch Eric Garcia

 rsbrennan-NOAA Reid Brennan

6009ec7 · 8 months ago

History

Raw

Preview Code Blame

⋮

Links and paths

NCBI

[https://github.com/ericgarciaresearch/noaa_sedna/
/share/all/ncbi_database/2025-04-01_ncbi
/share/all/REVAMP/2025-04-01_ncbi_forREVAMP](https://github.com/ericgarciaresearch/noaa_sedna/blob/main/ncbi/share/all/ncbi_database/2025-04-01_ncbi/share/all/REVAMP/2025-04-01_ncbi_forREVAMP)

MIDORI2

[https://github.com/ericgarciaresearch/noaa_sedna/
blob/main/midori2.md](https://github.com/ericgarciaresearch/noaa_sedna/blob/main/midori2.md)
/share/all/midori2_database

Key Take-Home Messages

- Reference databases define what you can detect
- Curation is essential for reliable eDNA inference
- SEDNA has **NCBI** (diff formats), **MIDORI2** (diff markers)
- MIDORI2 balances breadth, accuracy, and usability
- Database choice is a scientific decision —not a technical detail