SWEDISH METABARCODING NETWORK · HANDOUT

The ASV portal: from ASV tables to reusable biodiversity records

What it is, why it's worth the extra effort, and how to get your eDNA data in

What is the ASV portal?

The ASV portal of the Swedish Biodiversity Data Infrastructure (SBDI) is the place for **processed** metabarcoding data: your ASVs, their taxonomy and the sample metadata. It is the interface to DNA-based occurrence data in the SBDI Bioatlas and GBIF, where your detections become searchable biodiversity records next to all other species observations.

	ENA	ASV portal
Stores	Raw sequences	Processed ASVs with taxonomic annotation
Answers	What was sequenced?	What was found?
Reused as	Sequence data for re-analysis	Species occurrence records (Bioatlas, GBIF)

How it works

1. **Submit raw reads to ENA**
   see the SBDI guide to ENA submission (Webin), link below
2. **Denoise reads to ASVs**
   e.g. with the nf-core/ampliseq pipeline
3. **Fill in the upload template**
   download SBDI-ASV-template.xlsx; aligned with GBIF's Metabarcoding Data Toolkit (MDT)
4. **Upload in the portal**
   log in and request upload permission at asv-portal.biodiversitydata.se/submit
5. **Data-sharing contract & publication**
   SBDI contacts you for a data-sharing contract, then your data are published in the ASV database and Bioatlas

Get help – it's free

Video guide
Step-by-step walkthrough of the ASV portal upload
youtube.com/watch?v=a9ABLK0OzjU

NBIS data submission support
Free personal help with ENA submission, metadata and ASV upload
nbis.se/services/data-submission-support/apply

Guide to ENA submission (Webin)
biodiversitydata-se.github.io/mol-data/ena-metabar.html

SBDI support
Questions about the portal
docs.biodiversitydata.se/support

Why upload?

- ✓ **Visible & citable** – your dataset becomes a published, citable resource instead of a supplementary file
- ✓ **Always up to date** – ASVs are re-annotated as reference databases improve, without extra work for you
- ✓ **Findable as species records** – detections appear in the Bioatlas and flow on to GBIF
- ✓ **Comparable** – put your results next to other metabarcoding datasets and environmental layers
- ✓ **Meets requirements** – a concrete answer to funders' FAIR and data management plan demands

ENA and ASV portal: you need both
ENA stores the raw sequences, the ASV portal the processed ASVs along with their taxonomic annotation and the place and time of sampling. Only together are your data fully reusable, linked via your ENA accessions.

Before you start, have ready

- ✓ ASV table with read counts per sample
- ✓ ASV sequences and taxonomic assignments
- ✓ Sample metadata: date, coordinates, sampling protocol
- ✓ ENA accessions for samples and sequence runs
- ✓ Methods: target gene & region, primers, sequencing platform, denoising tool

Already submitted to ENA? Then you have most of the sample metadata ready, and your ENA accessions link the two records.

Swedish Metabarcoding Network
metabarcoding.se · smn-members@lists.uu.se
Research Data Community Project funded by SciLifeLab

metabarcoding.se

What goes into the template: four sheets

Example: **LMO 16S rRNA demo dataset** (Linnaeus Microbial Observatory, Baltic Sea, 16S V3–V4, processed with nf-core/ampliseq), as shown in the live demo. Shown here: a subset of 4 samples and 5 ASVs. Download the empty template at asv-portal.biodiversitydata.se/files/SBDI-ASV-template.xlsx.

1 · ASV-table **Abundances** – one row per ASV, one column per sample, cells = read counts

asv_id	P2256_1001	P1313_2035	P8309_1088	P31912_1090
1ead9875…	1 270	5 340	345	1 858
487ba8cf…	305	923	326	27
f2afda01…	191	72	34	342
e81e42b8…	40	82	37	44
28894e2c…	61	34	50	13

2 · Taxonomy **Sequences & taxonomy** – the same asv_ids, with their DNA sequence and your taxonomic assignment

asv_id	DNA_sequence	phylum	class	order	family	genus	specificEpithet
1ead9875…	TGGGGAATTTTCCG…	Cyanobacteriota	Cyanobacteriia	PCC-6307	Cyanobiaceae	Cyanobium_A	usitatum
487ba8cf…	TCGAGAATCTTCCG…	Planctomycetota	Planctomycetia	Planctomycetales	Planctomycetaceae	Fuerstiella	marisgermanici
f2afda01…	TGGGGAATCTTGCA…	Pseudomonadota	Alphaproteobacteria	Pelagibacterales	Pelagibacteraceae	Pelagibacter	ubique
e81e42b8…	TGAGGAATATTGGG…	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Polaribacter	gangjinensis
28894e2c…	TGAGGAATATTGGA…	Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Ulvibacter	antarcticus

Also: kingdom, infraspecificEpithet, otu. The portal later re-annotates the ASVs against current reference databases.

3 · Samples **Sample metadata** – one row per sample; sample_id matches the column names in the ASV-table

sample_id	materialSample ID	associated Sequences	eventDate	decimal Latitude	decimal Longitude	maximumDepthIn Meters	salinity (psu)	temperature (°C)
P2256_1001	SAMEA14374254	ERR9715422	2011-03-25	56.9309	17.0607	2	7.0	1.5
P1313_2035	SAMEA7871181	ERR5191477	2013-11-26	56.9309	17.0607	2	6.9	6.8
P8309_1088	SAMEA14375002	ERR9717165	2017-04-27	56.9309	17.0607	2	6.6	4.7
P31912_1090	SAMEA117293515	ERR13966845	2023-10-03	56.9309	17.0607	2	7.3	12.2

materialSampleID and associatedSequences link to the ENA sample and run (shortened here to the accession). Also: samplingProtocol, locationID, geodeticDatum, country, verbatimLocality, filter fraction, …

4 · Study **Study information** – filled in once per dataset

datasetName	LMO 16S rRNA metabarcoding dataset (demo subset)
institutionCode	LNU
target_gene	16S rRNA
target_subfragment	V3-V4
pcr_primer_forward	CCTACGGGNGGCWGCAG
pcr_primer_reverse	GGACTACHVGGGTWTCTAAT
seq_meth	Illumina MiSeq
denoising_appr	DADA2
sop	https://nf-co.re/ampliseq

Further fields: datasetID, institutionID (ROR), primer names, library layout and environment terms from ENVO (broad scale, local scale, medium).