The Swedish ASV Portal for Molecular Biodiversity Data

Making eDNA data open and FAIR

Rasa Bukontaite

SBDI coordinator
Swedish Museum of Natural History

Tobias Andermann Maria Prager Anders Andersson Daniel Lundin

SBDI

GBIF NODE

SWEDISH MUSEUM OF NATURAL HISTORY

SLU

UNIVERSITY OF GOTHENBURG

Karolinska Institutet

Linnæus University

LUND UNIVERSITY

KTH VETENSKAP OCH KONST

SMHI

Stockholm University

UMEÅ UNIVERSITET

UPPSALA UNIVERSITET

Swedish Research Council

SBDI

SWEDISH BIODIVERSITY
DATA INFRASTRUCTURE

Open Science

Open Data

FAIR & CARE

SBDI is developed in collaboration with the Living Atlases community, and includes the Swedish node of the Global Biodiversity Information Facility (GBIF).

SBDI: ~168 million Occurrence Records
GBIF: >3.5 billion Occurrence Records

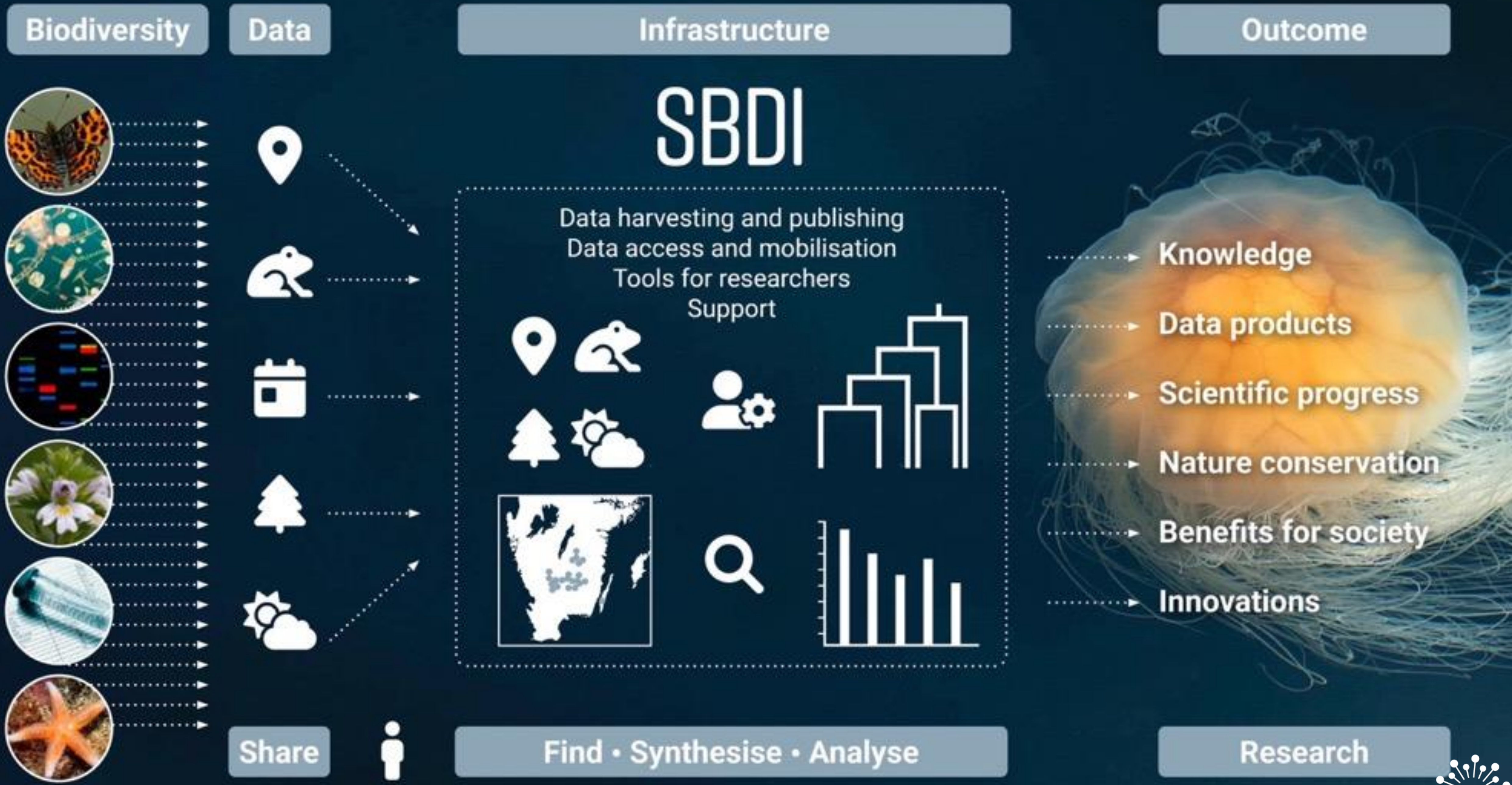
Living Atlases

An open community created around the Atlas of Living Australia platform.

GBIF Data from Sweden

Swedish publishing trends

Search data published by Sweden



Biodiversity **Data**

Infrastructure

SBDI

Data harvesting and publishing
Data access and mobilisation
Tools for researchers
Support

Outcome

- Knowledge
- Data products
- Scientific progress
- Nature conservation
- Benefits for society
- Innovations

Share

Find • Synthesise • Analyse

Research

www.biodiversitydata.se

Access to Biodiversity Data and Tools

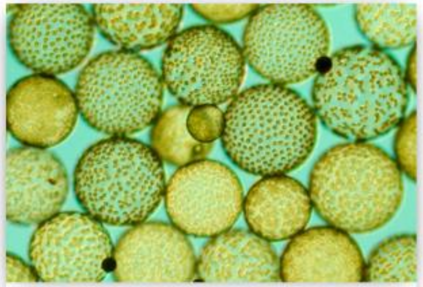
www.biodiversitydata.se

Swedish Biodiversity Data

Explore by Location

Searching and Mapping

R Tools and APIs



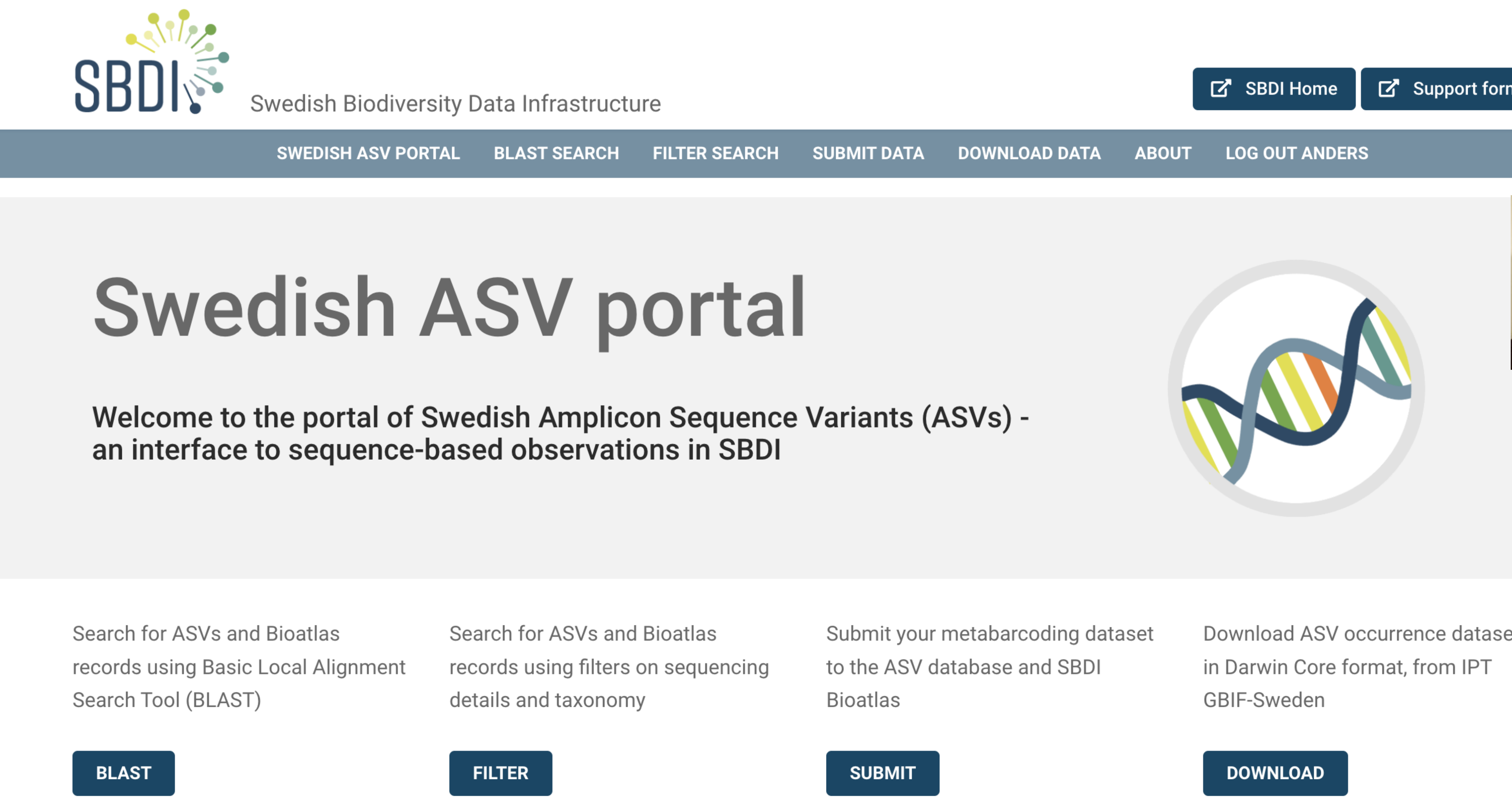
Molecular Biodiversity Data

Systematic Monitoring and Citizen Science Tools

Marine Data Tools

Biologging Data Tools

Archaeological Biodiversity Data Tools



Swedish Biodiversity Data Infrastructure

Swedish ASV portal

Welcome to the portal of Swedish Amplicon Sequence Variants (ASVs) - an interface to sequence-based observations in SBDI

Search for ASVs and Bioatlas records using Basic Local Alignment Search Tool (BLAST)

Search for ASVs and Bioatlas records using filters on sequencing details and taxonomy

Submit your metabarcoding dataset to the ASV database and SBDI Bioatlas

Download ASV occurrence datasets, in Darwin Core format, from IPT GBIF-Sweden

Maria Prager

asv-portal.biodiversitydata.se

Stockholm University
UPPSALA UNIVERSITET
UNIVERSITY OF GOTHENBURG
Linnæus University

Swedish ASV
Biodiversity Metabarcoding Portal

A dedicated interface for accessing and analyzing sequence-based biodiversity observations through SBDI

- **Curated database** with user-friendly web interface
- **Automated analysis tools:** nf-core pipeline for denoising & taxonomic annotation
- **Submission support:** streamlined workflow for depositing raw data to ENA

Recently Published DNA-derived Datasets from Sweden

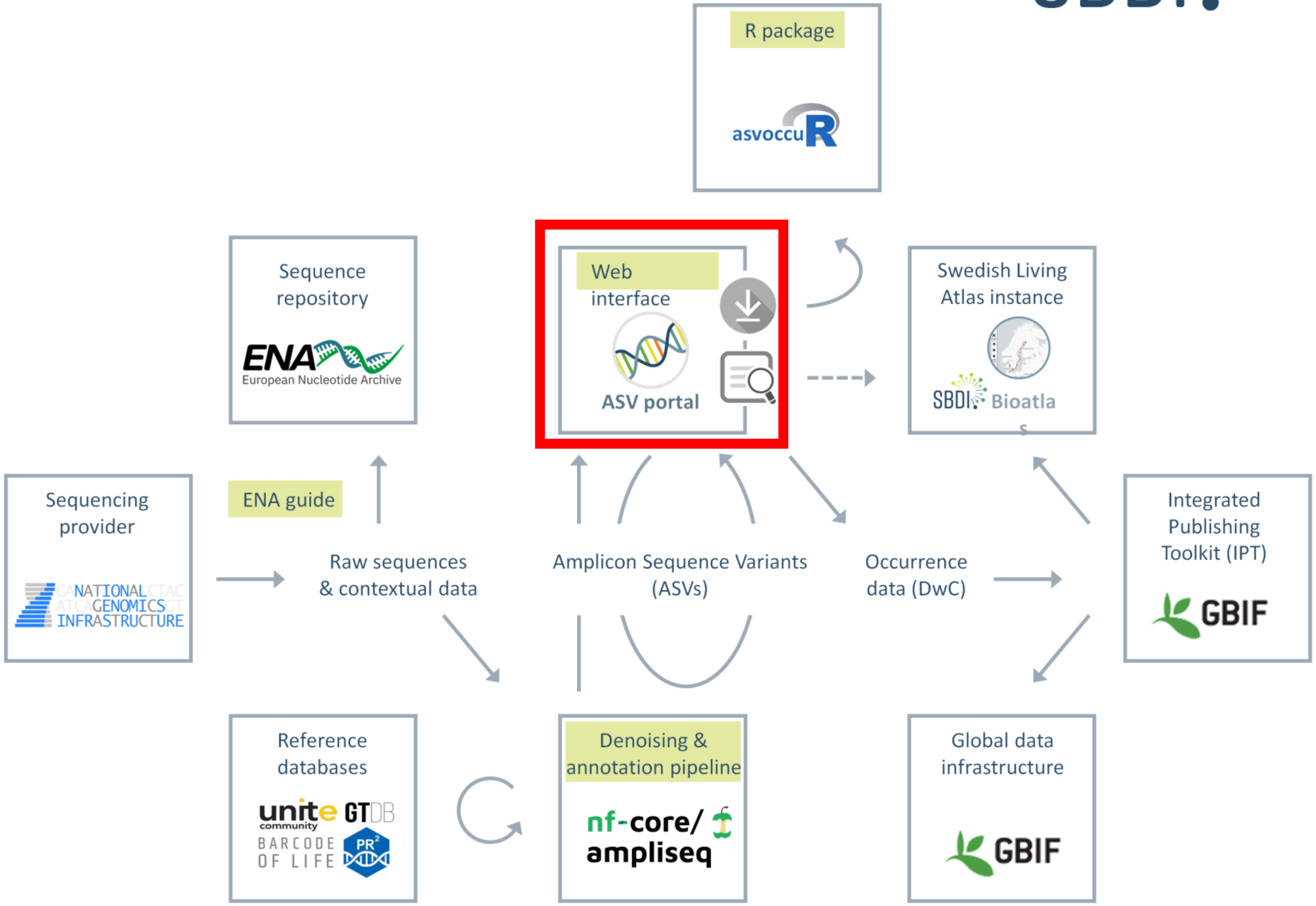
21 datasets have been published so far!

- CO1 Amplicon Sequence Variants of soil and leaf litter arthropod communities collected at Malaise traps from the Insect Biome Atlas project in Sweden
- CO1 Amplicon Sequence Variants of bulk arthropod samples (mild lysis) collected with Malaise traps from the Insect Biome Atlas project in Sweden
- Fungal Diversity Survey Sweden ITS-LSU rDNA
- COI data from: Invasive species detection along coastal harbours in northern region of Vastra Gotaland 2024
- COI Metabarcoding Genetic Observations of Marine Species in the Port of Wallhamn, Sweden (2022)
- 18S Metabarcoding Genetic Observations of Marine Species in the Port of Wallhamn, Sweden (2022)
- 16S data from: Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea (Hu et al. 2016)
- Kungsängen soil microbial communities
- 18S data from: Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea (Hu et al. 2016)
- Environmental long read amplicons of soil fungi across Podzol soil profile
- Soil fungal communities of ectomycorrhizal dominated woodlands across West Africa

Veronika A. Johansson

Swedish ASV
- publishing and accessing metabarcoding data

- **Sequencing**
  - A service provider (e.g. **NGI**) delivers sequencing data to the user.
- **Submission to ENA**
  - The user uploads **raw sequencing data and contextual metadata** to the **ENA** *(see our guide)*.
- **Data Processing**
  - The user **denoises the data** using tools such as **nf-core/ampliseq**.
- **Submission to ASV Portal**
  - **Denoised data** are submitted to the **SBDI ASV Portal** for integration and access.
- **Data Access & Use**
  - **Metabarcoding data** can be **searched and downloaded** via the ASV Portal.
  - **Downloaded datasets** can be further analyzed using the **asvoccur R package**.



R package

asvoccuR

Sequence repository

ENA European Nucleotide Archive

Web interface

ASV portal

Swedish Living Atlas instance

SBDI Bioatlas

Sequencing provider

NATIONAL GENOMICS INFRASTRUCTURE

ENA guide

Raw sequences & contextual data

Amplicon Sequence Variants (ASVs)

Occurrence data (DwC)

Integrated Publishing Toolkit (IPT)

GBIF

Reference databases

unite community GTDB BARCODE OF LIFE PR²

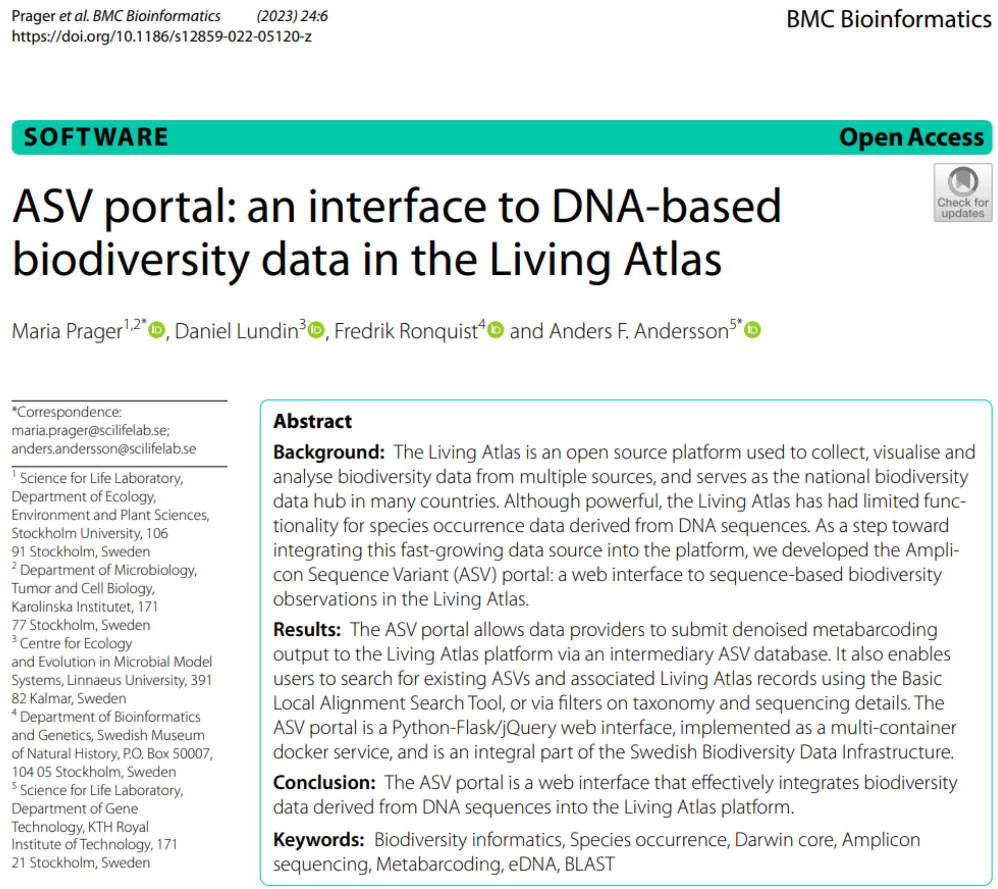
Denoising & annotation pipeline

nf-core/ ampliseq

Global data infrastructure

GBIF

Swedish ASV structure - technical overview



Prager *et al. BMC Bioinformatics* (2023) 24:6
https://doi.org/10.1186/s12859-022-05120-z

BMC Bioinformatics

SOFTWARE **Open Access**

ASV portal: an interface to DNA-based biodiversity data in the Living Atlas

Maria Prager[1,2*], Daniel Lundin[3], Fredrik Ronquist[4] and Anders F. Andersson[5*]

*Correspondence:
maria.prager@scilifelab.se;
anders.andersson@scilifelab.se

[1] Science for Life Laboratory, Department of Ecology, Environment and Plant Sciences, Stockholm University, 106 91 Stockholm, Sweden
[2] Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 171 77 Stockholm, Sweden
[3] Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, 391 82 Kalmar, Sweden
[4] Department of Bioinformatics and Genetics, Swedish Museum of Natural History, P.O. Box 50007, 104 05 Stockholm, Sweden
[5] Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, 171 21 Stockholm, Sweden

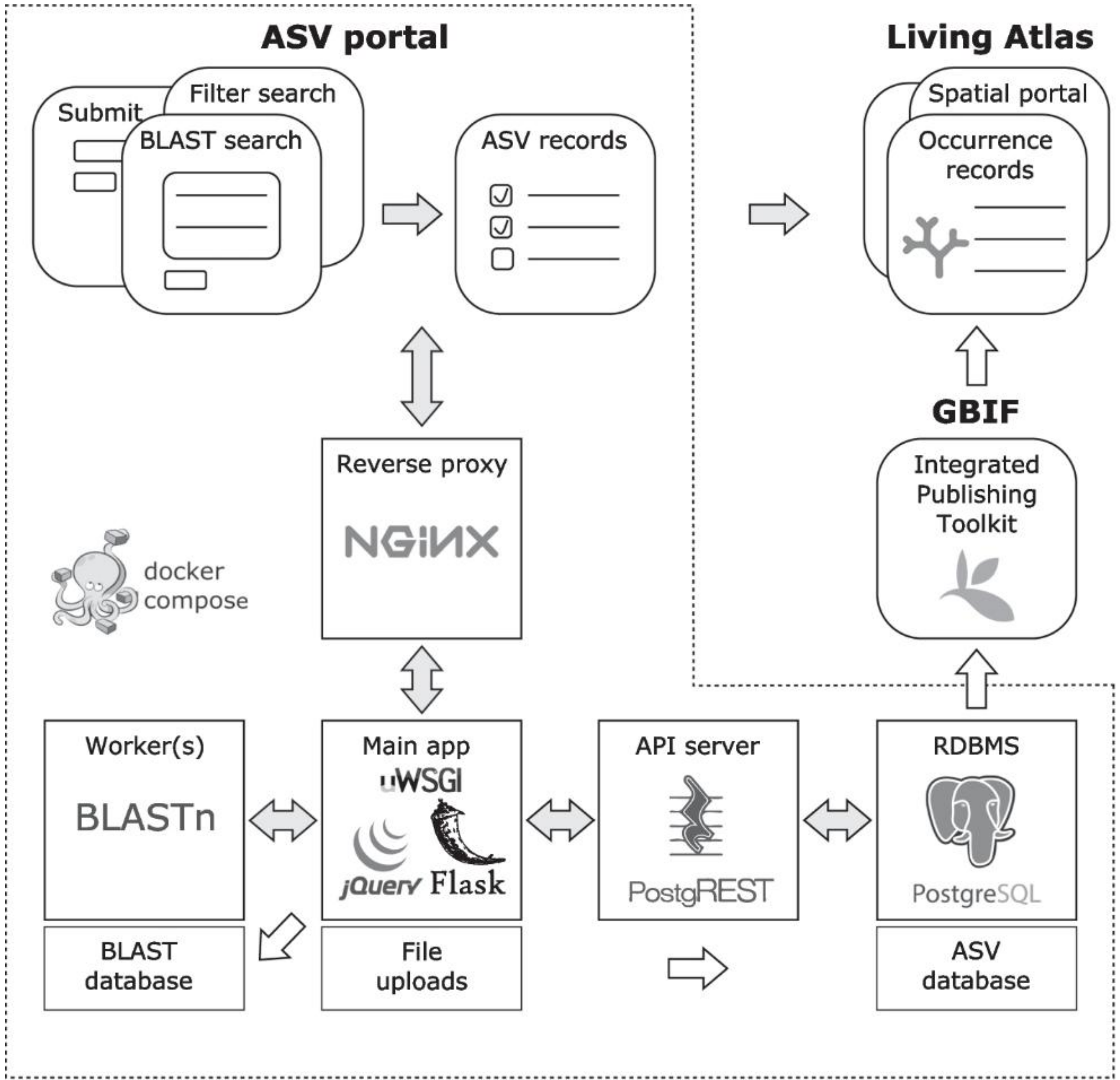
Abstract

Background: The Living Atlas is an open source platform used to collect, visualise and analyse biodiversity data from multiple sources, and serves as the national biodiversity data hub in many countries. Although powerful, the Living Atlas has had limited functionality for species occurrence data derived from DNA sequences. As a step toward integrating this fast-growing data source into the platform, we developed the Amplicon Sequence Variant (ASV) portal: a web interface to sequence-based biodiversity observations in the Living Atlas.

Results: The ASV portal allows data providers to submit denoised metabarcoding output to the Living Atlas platform via an intermediary ASV database. It also enables users to search for existing ASVs and associated Living Atlas records using the Basic Local Alignment Search Tool, or via filters on taxonomy and sequencing details. The ASV portal is a Python-Flask/jQuery web interface, implemented as a multi-container docker service, and is an integral part of the Swedish Biodiversity Data Infrastructure.

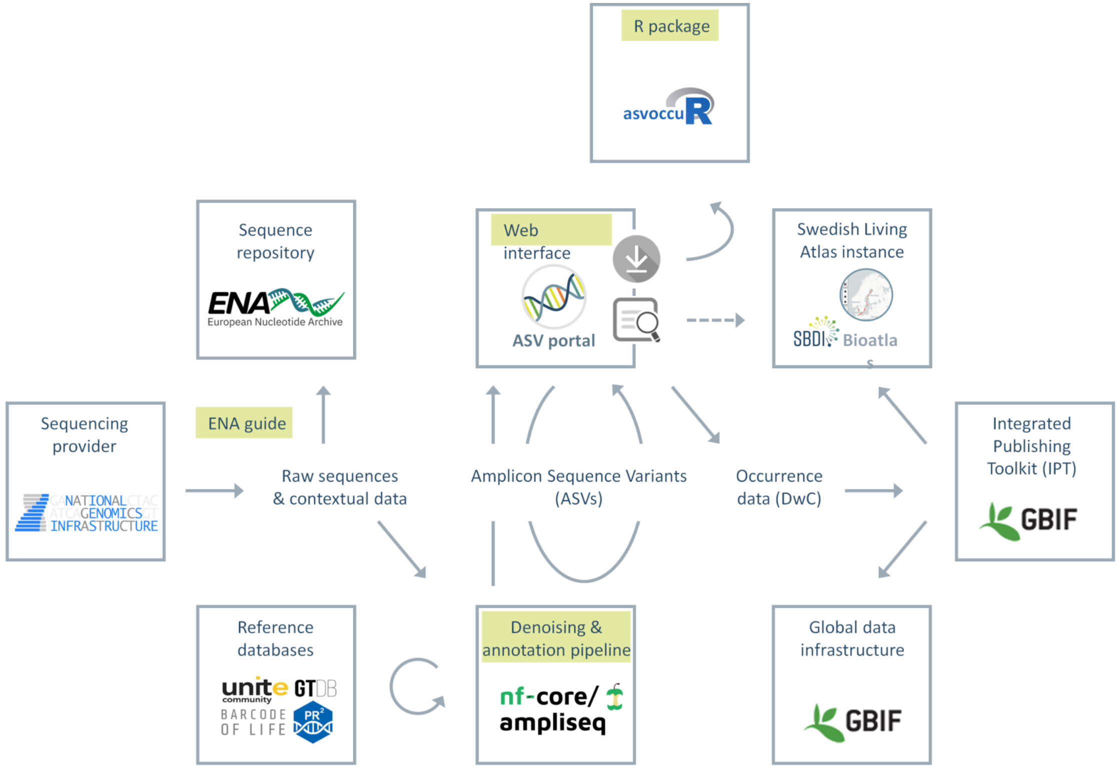
Conclusion: The ASV portal is a web interface that effectively integrates biodiversity data derived from DNA sequences into the Living Atlas platform.

Keywords: Biodiversity informatics, Species occurrence, Darwin core, Amplicon sequencing, Metabarcoding, eDNA, BLAST



Swedish ASV - highlights

- **ASV storage** and dynamic approach to **taxonomic annotation**
- **BLAST** search
- R package: **asvoccur**



R package
asvoccuR

Sequence repository
ENA European Nucleotide Archive

Web interface
ASV portal

Swedish Living Atlas instance
SBDI Bioatlas

Sequencing provider
NATIONAL GENOMICS INFRASTRUCTURE

ENA guide

Raw sequences & contextual data

Amplicon Sequence Variants (ASVs)

Occurrence data (DwC)

Integrated Publishing Toolkit (IPT)
GBIF

Reference databases
unite GTDB BARCODE OF LIFE PR2

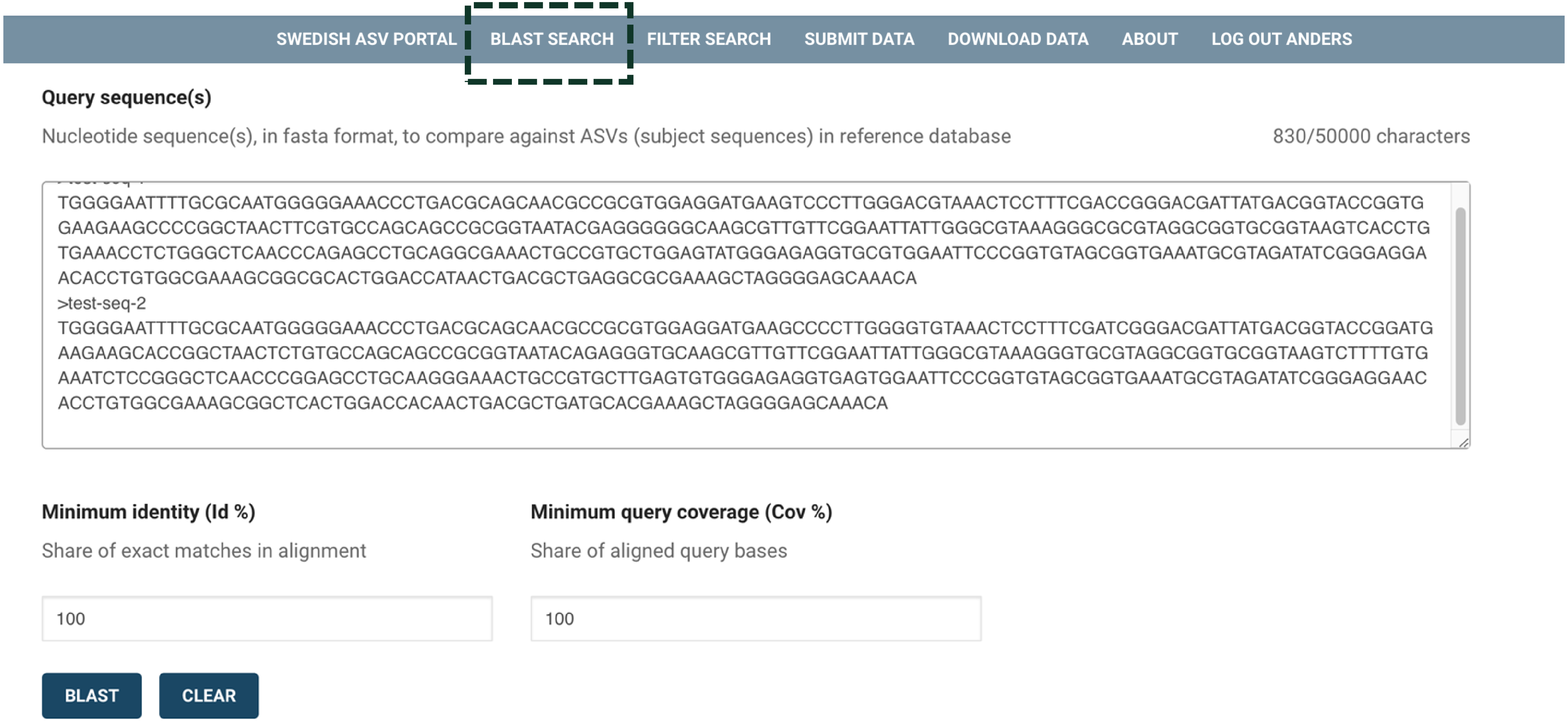
Denoising & annotation pipeline
nf-core/ ampliseq

Global data infrastructure
GBIF

Swedish ASV- storage and taxonomic re-annotation

- Submitted data files are curated and important into **ASV database**
- A **standard taxonomic annotation** is applied to each ASV using up-to-date **classification algorithms** and **reference databases**.
- It allows for successive re-annotations, enabling improved taxonomic accuracy and resolution as reference database develop.
- The **database structure** supports **successive re-annotations**, improving taxonomic accuracy and resolution as reference databases evolve.
- Each **DwC (Darwin Core) occurrence** is assigned a **unique taxon ID** for traceability and consistency.

Swedish ASV- BLAST search



Query sequence(s)

Nucleotide sequence(s), in fasta format, to compare against ASVs (subject sequences) in reference database

830/50000 characters

```
TGGGGAATTTTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGGAGGATGAAGTCCCTTGGGACGTAAACTCCTTTCGACCGGGACGATTATGACGGTACCGGTG
GAAGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAGGGGGCAAGCGTTGTTCGGAATTATTGGGCGTAAAGGGCGCGTAGGCGGTGCGGTAAGTCACCTG
TGAAACCTCTGGGCTCAACCCAGAGCCTGCAGGCGAAACTGCCGTGCTGGAGTATGGGAGAGGTGCGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGA
ACACCTGTGGCGAAAGCGGCGCACTGGACCATAACTGACGCTGAGGCGCGAAAGCTAGGGGAGCAAACA
>test-seq-2
TGGGGAATTTTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGGAGGATGAAGCCCCTTGGGGTGTAAACTCCTTTCGATCGGGACGATTATGACGGTACCGGATG
AAGAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTGTTCGGAATTATTGGGCGTAAAGGGTGCGTAGGCGGTGCGGTAAGTCTTTTGTG
AAATCTCCGGGCTCAACCCGGAGCCTGCAAGGGAAACTGCCGTGCTTGAGTGTGGGAGAGGTGAGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAAC
ACCTGTGGCGAAAGCGGCTCACTGGACCACAACTGACGCTGATGCACGAAAGCTAGGGGAGCAAACA
```

Minimum identity (Id %)

Share of exact matches in alignment

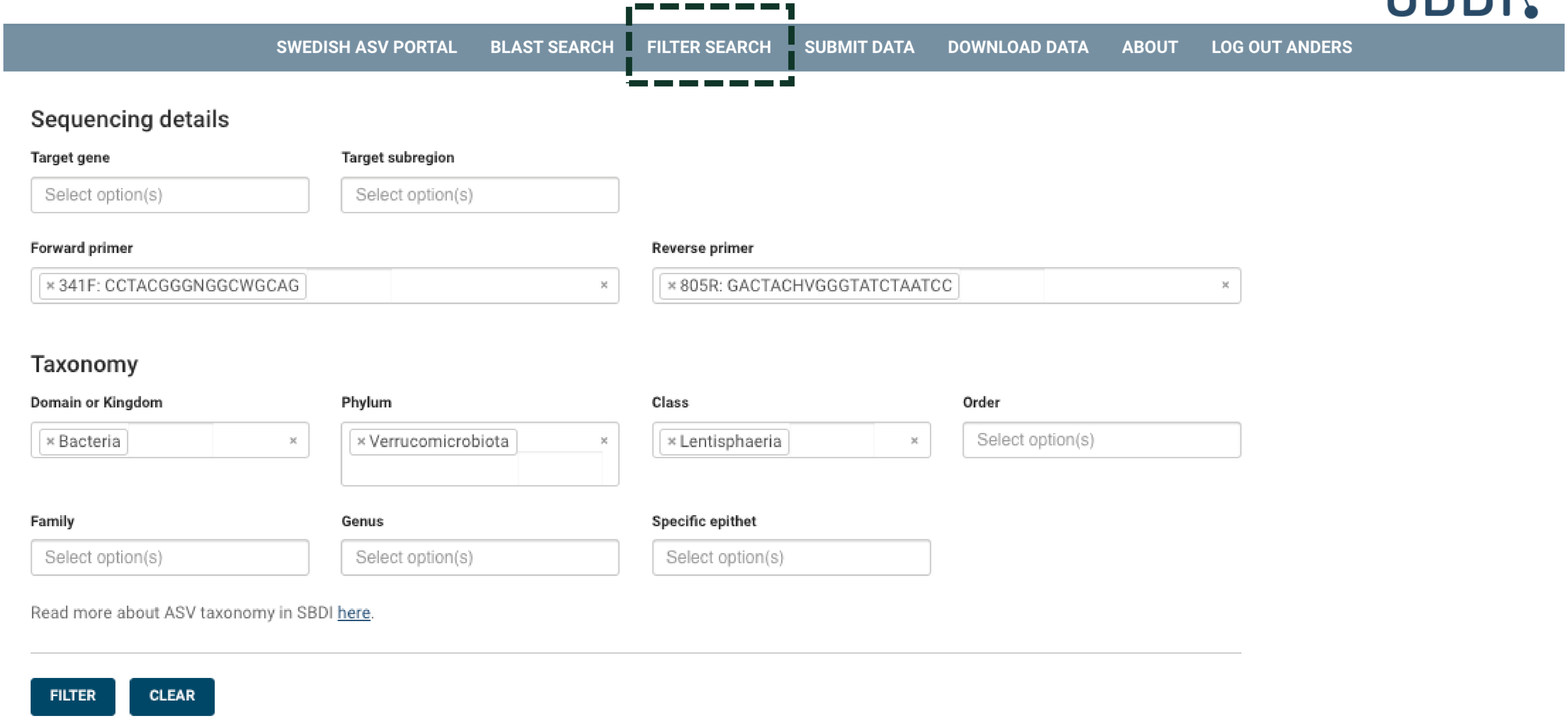
100

Minimum query coverage (Cov %)

Share of aligned query bases

100

Swedish ASV- BLAST search



Sequencing details

Target gene: Select option(s)

Target subregion: Select option(s)

Forward primer: × 341F: CCTACGGGNGGCWGCAG

Reverse primer: × 805R: GACTACHVGGGTATCTAATCC

Taxonomy

Domain or Kingdom: × Bacteria

Phylum: × Verrucomicrobiota

Class: × Lentisphaeria

Order: Select option(s)

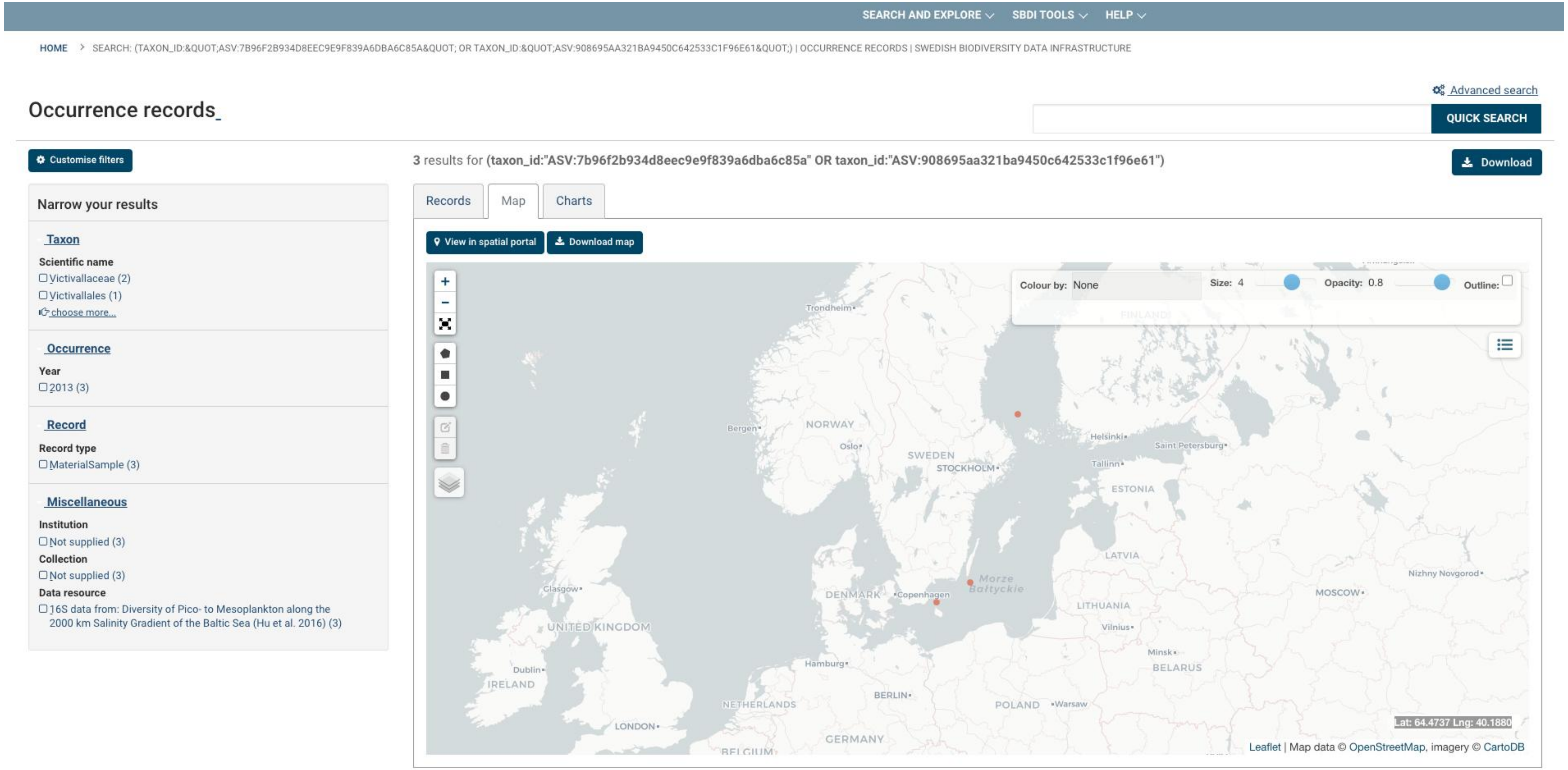
Family: Select option(s)

Genus: Select option(s)

Specific epithet: Select option(s)

Read more about ASV taxonomy in SBDI here.

Swedish ASV- BLAST search



HOME > SEARCH: (TAXON_ID:"ASV:7B96F2B934D8EEC9E9F839A6DBA6C85A" OR TAXON_ID:"ASV:908695AA321BA9450C642533C1F96E61") | OCCURRENCE RECORDS | SWEDISH BIODIVERSITY DATA INFRASTRUCTURE

Occurrence records_

3 results for **(taxon_id:"ASV:7b96f2b934d8eec9e9f839a6dba6c85a" OR taxon_id:"ASV:908695aa321ba9450c642533c1f96e61")**

Narrow your results

Taxon

Scientific name
☐ Victivallaceae (2)
☐ Victivallales (1)
choose more...

Occurrence

Year
☐ 2013 (3)

Record

Record type
☐ MaterialSample (3)

Miscellaneous

Institution
☐ Not supplied (3)

Collection
☐ Not supplied (3)

Data resource
☐ 16S data from: Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea (Hu et al. 2016) (3)

Colour by: None Size: 4 Opacity: 0.8 Outline: ☐

Lat: 64.4737 Lng: 40.1880

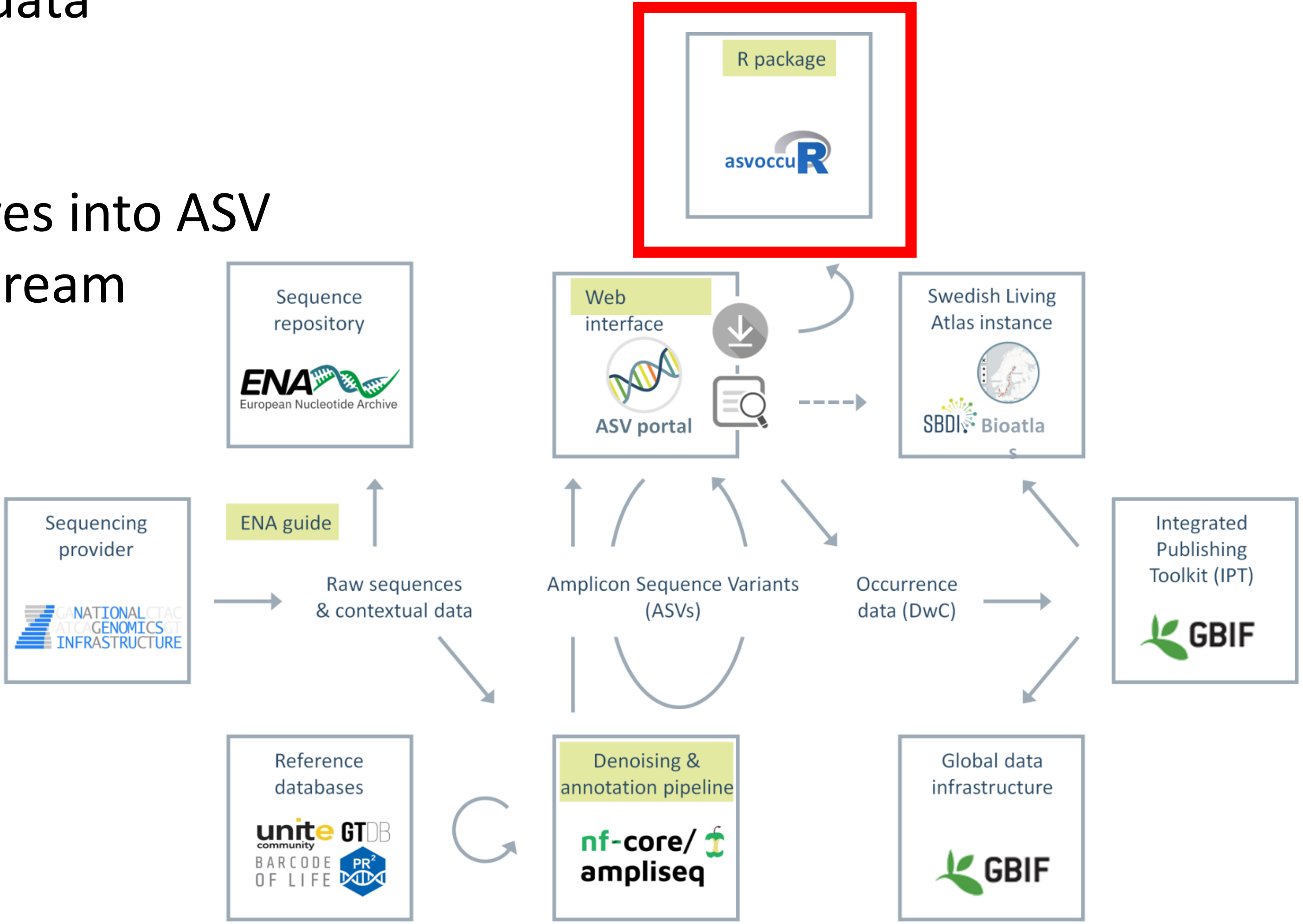
Swedish ASV- R package: asvoccur

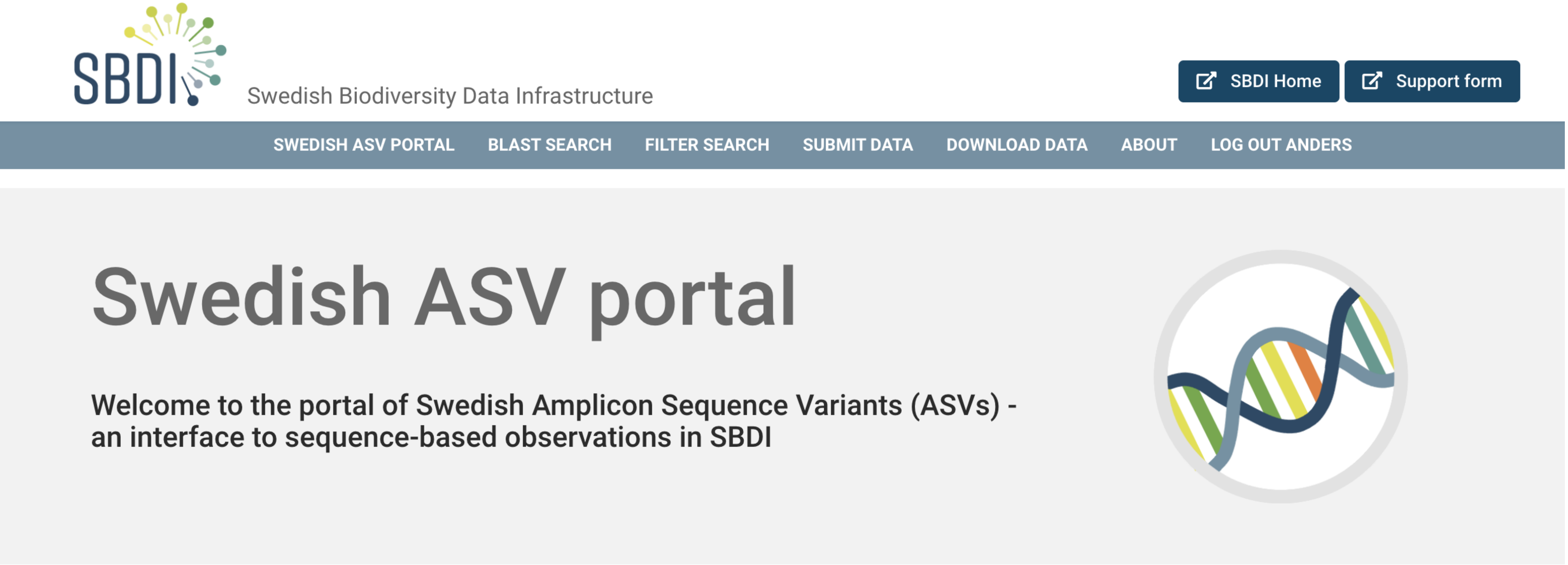
R package for processing occurrence data downloaded from the ASV portal.

It converts the condensed DwC archives into ASV table format, which facilitates downstream analysis in R.

The package includes functions:

- loading and merging datasets
- aggregating ASV counts across taxonomic ranks





Swedish ASV portal

Welcome to the portal of Swedish Amplicon Sequence Variants (ASVs) - an interface to sequence-based observations in SBDI

Any questtions?

Contact support: https://docs.biodiversitydata.se/support

SBDI

SWEDISH BIODIVERSITY
DATA INFRASTRUCTURE

Thank you!

A heartfelt thank you to the ***ASV Working Group*** *— your dedication and expertise have been the driving force behind SBDI's progress in the metabarcoding world.*

We wouldn't be here without you!

www.biodiversitydata.sec