SBDI Swedish Biodiversity Data Infrastructure

Make your data count: Process, publish and reuse your metabarcoding data in SBDI

Anders Andersson, KTH/SciLifeLab
Maria Prager, SU/KI
Jeanette Tångrot, UmU/NBIS

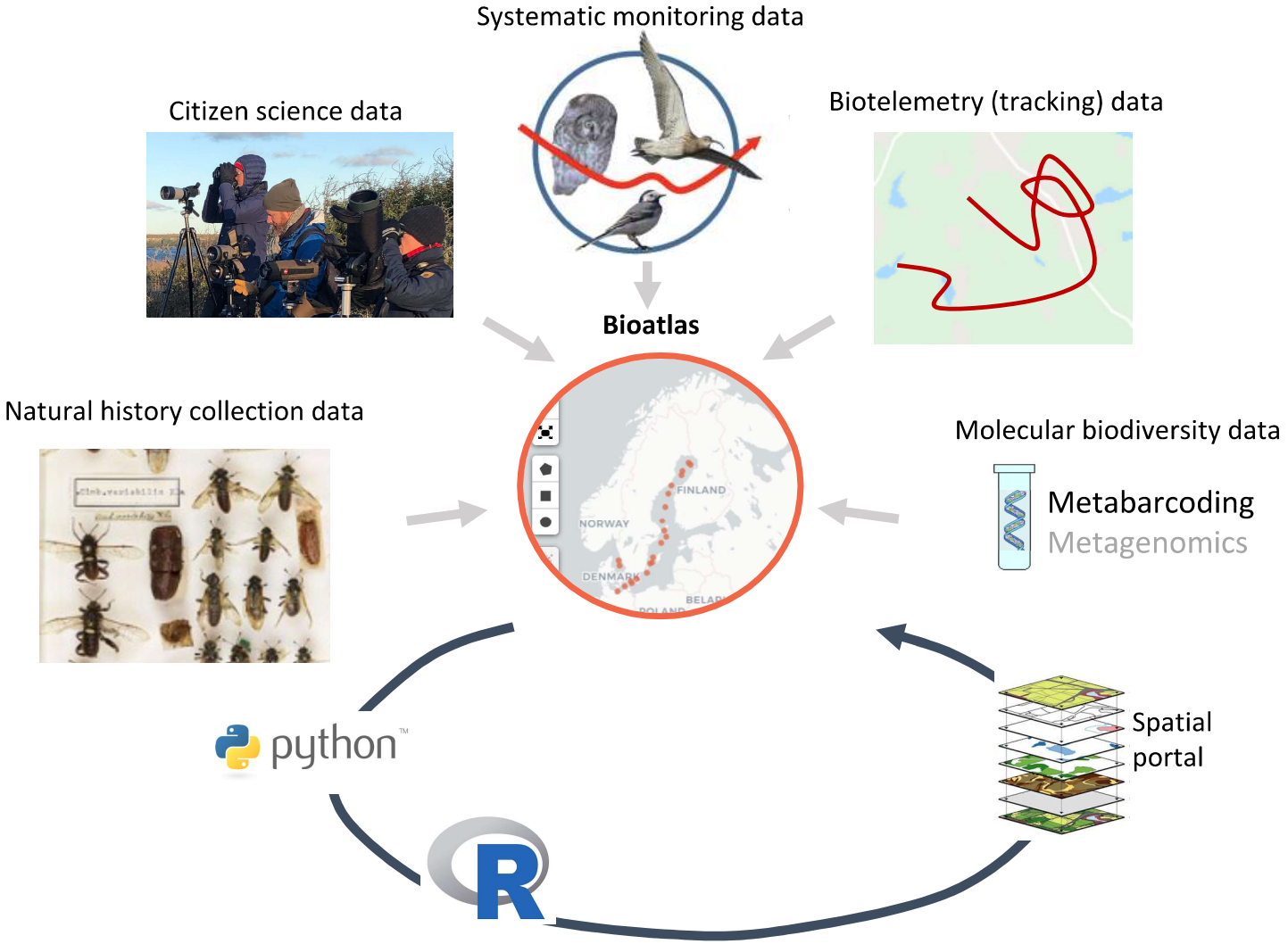
Outline

- Introduction to SBDI, GBIF and genetic data in SBDI
- Publishing metabarcoding data in ENA
- The nf-core/ampliseq pipeline for metabarcoding data
- Publishing, searching and downloading metabarcoding data in SBDI

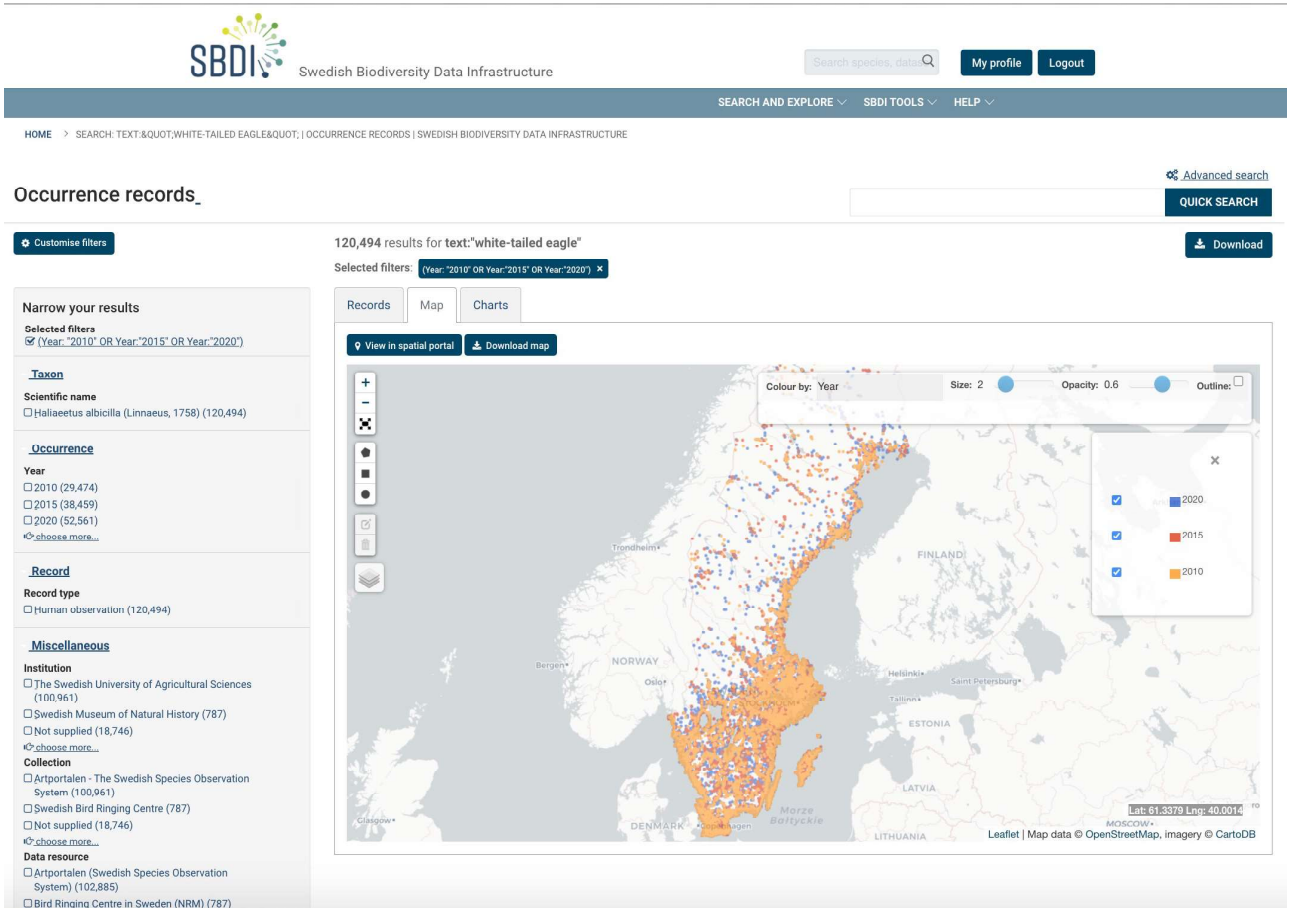
SBDI – Swedish Biodiversity Data Infrastructure

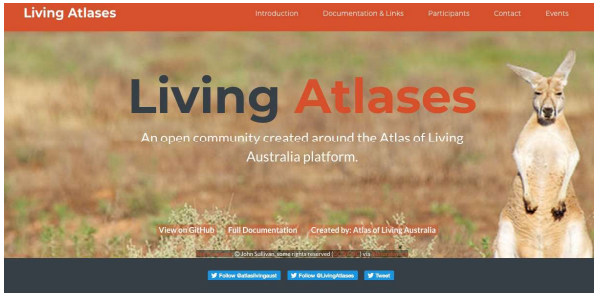
- National research infrastructure supported by the Swedish Research Council (VR)
- 11 universities and government agencies
- Host: Swedish Museum of Natural History
- Swedish node of GBIF
- Focused on mobilizing biodiversity data and providing tools for analysis and visualization
  - Citizen science data (from bird watchers, amateur botanists, etc.)
  - Data from systematic monitoring programs
  - Data from natural history museums
  - Genetic data from DNA sequencing of environmental samples (eDNA)
  - Etc.

The SBDI Bioatlas



The SBDI Bioatlas

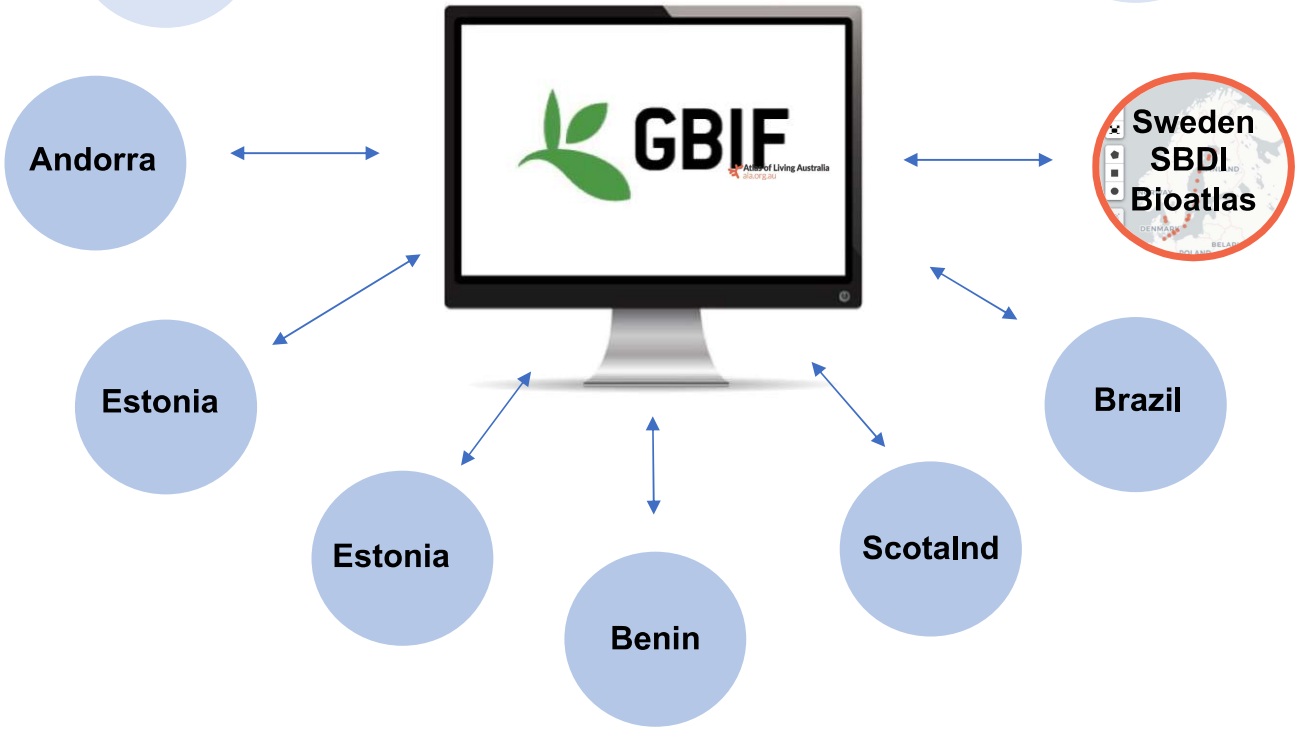




- Built on the Atlas of Living Australia (launched 2010)
- 19 National Atlases currently set up
- Open-source software
- Storage, mobilisation, analysis, visualisation of biodiversity data
- Analysis platform, R tools, web services

GBIF – The Global Biodiversity Information Facility

- **GBIF:** A global network and data infrastructure providing open access to biodiversity data.
- **Data Scope:** Covers species occurrences from museum specimens to citizen science, across all life forms.
- **Mission:** Supports research, conservation, and environmental policy.
- **Collaborative effort**: Backed by governments and organizations worldwide, pooling diverse data sources.
- **Impact:** Aids studies in ecology, conservation, and climate change by offering essential data to researchers and policymakers.



GBIF – The Global Biodiversity Information Facility

Datasets
108,405

Hosted portals
23

Country Participants
63

Peer-review papers using data
11,034

Organizational Participants
43

Average records downloaded per month (2024)
201.5 billion

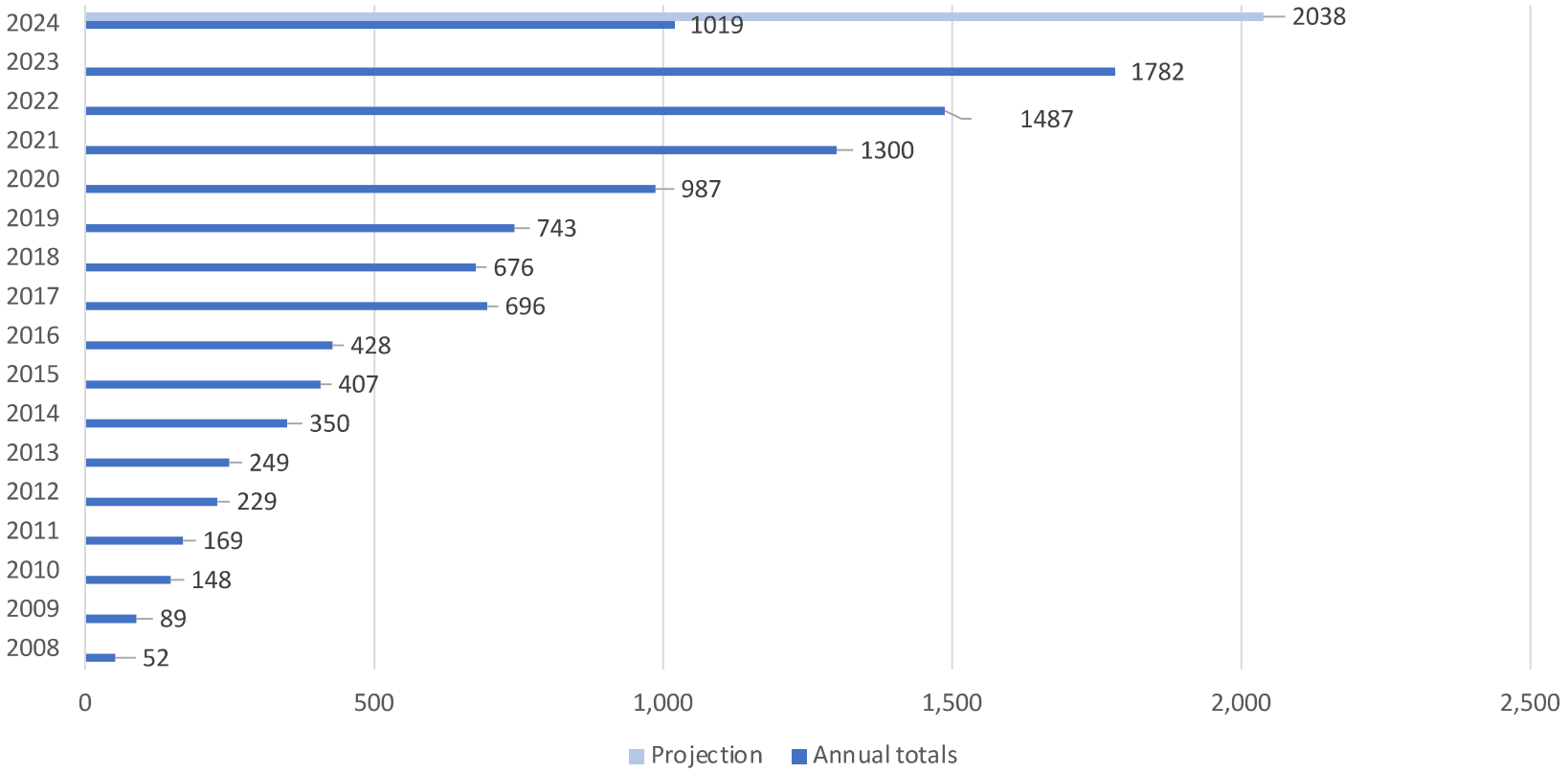
Publishers
2,282

Species occurrence records
3,005,160,507

By the numbers | 13 September 2024

www.gbif.org

Peer-reviewed publications using GBIF data



30 June 2024

https://www.gbif.org/resource/search?contentType=literature&literatureType=journal&relevance=GBIF_USED&peerReview=true

Examples of studies using GBIF data (out of >11,000)

Exponential increase in records of invasive alien species worldwide

Mormul, *Biol Invasions*, 2022

Mapping the global impact of agriculture on biodiversity

Hoang et al. *PNAS*, 2023

Insights and gaps: mapping the diversity of Europe's wild bees

Leclercq, *Journal of Biogeography*, 2023

Assessing impact of stressors on global penguin hotspots

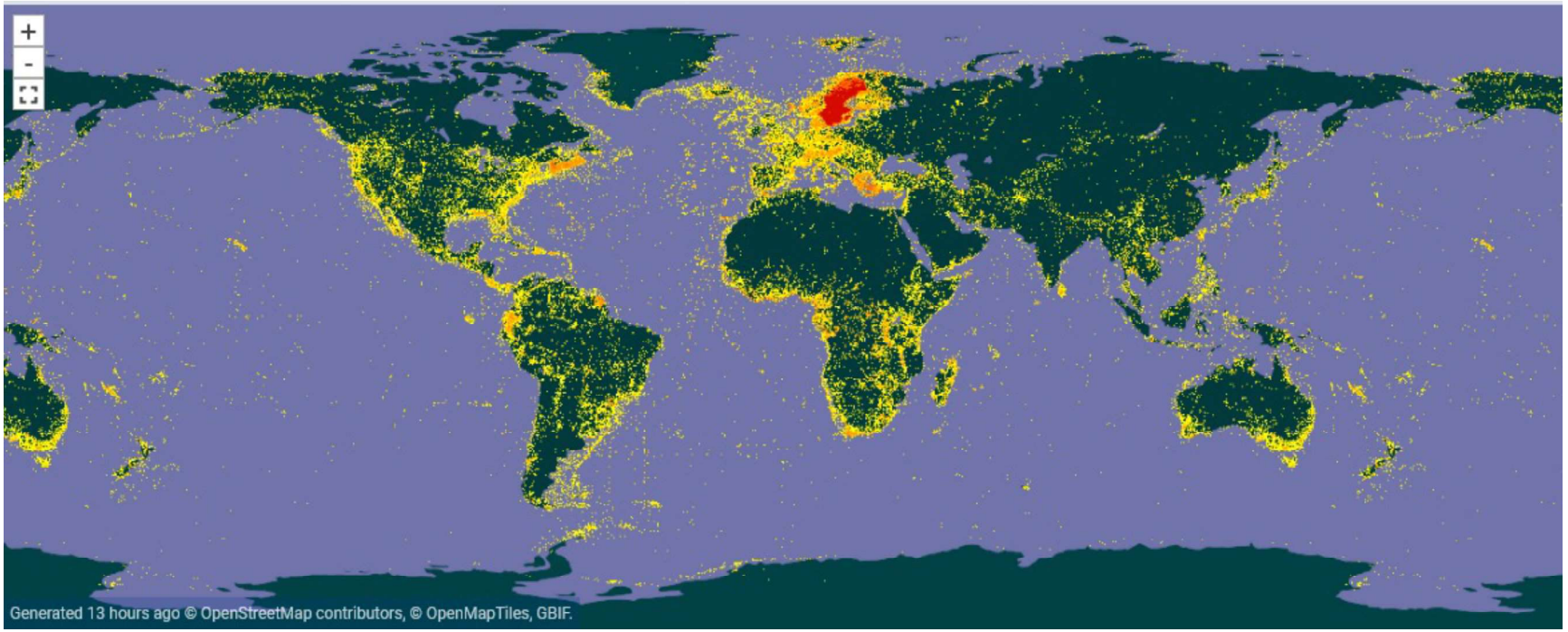
Gimeno et al. *Global Change Biology*, 2024

https://www.gbif.org/document/5N9YVBkTP3y7kqhFQviowM/gbif-science-review-no-11

GBIF Sweden – data mobilization so far

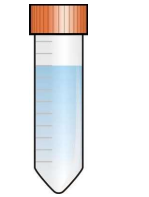
DATA FROM SWEDEN

135,573,631	160	248	28
Published occurrences	Published datasets	Countries and areas covered by data from Sweden	Publishers from Sweden



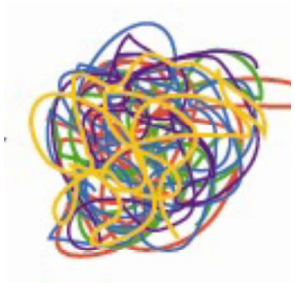
https://www.gbif.org/resource/search?contentType=literature&literatureType=journal&relevance=GBIF_USED&peerReview=true

Metabarcoding



Sample
(water, soil, insect trap, …)

DNA extraction



DNA

PCR

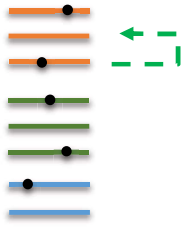
PCR-amplified taxonomic marker genes

DNA sequencing

```
>seq1
AATGTAGTA…
>seq2
AACGTAGGT…
>seq3
AACCCTTAC…
```

sequences (barcodes)

Error correction (denoising)



Amplicon Sequence Variants (ASVs)

Taxonomic annotation

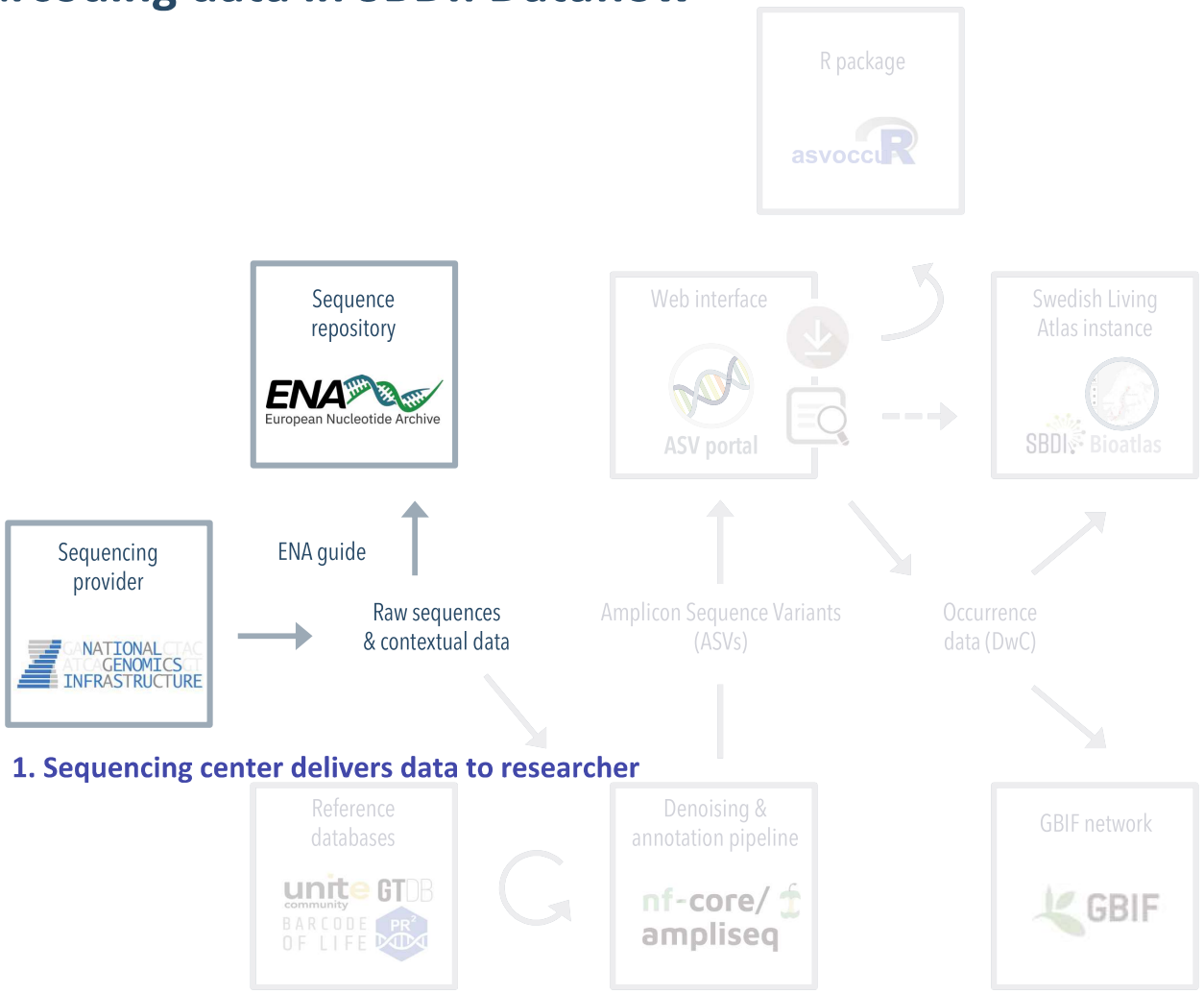
(using reference sequence database, e.g. GTDB, UNITE, BOLD)

ASV	Species	Sample1	Sample2
1	E.coli	17	0
2	S.aurus	231	11800
3	unknown	30	0
…	…	…	…

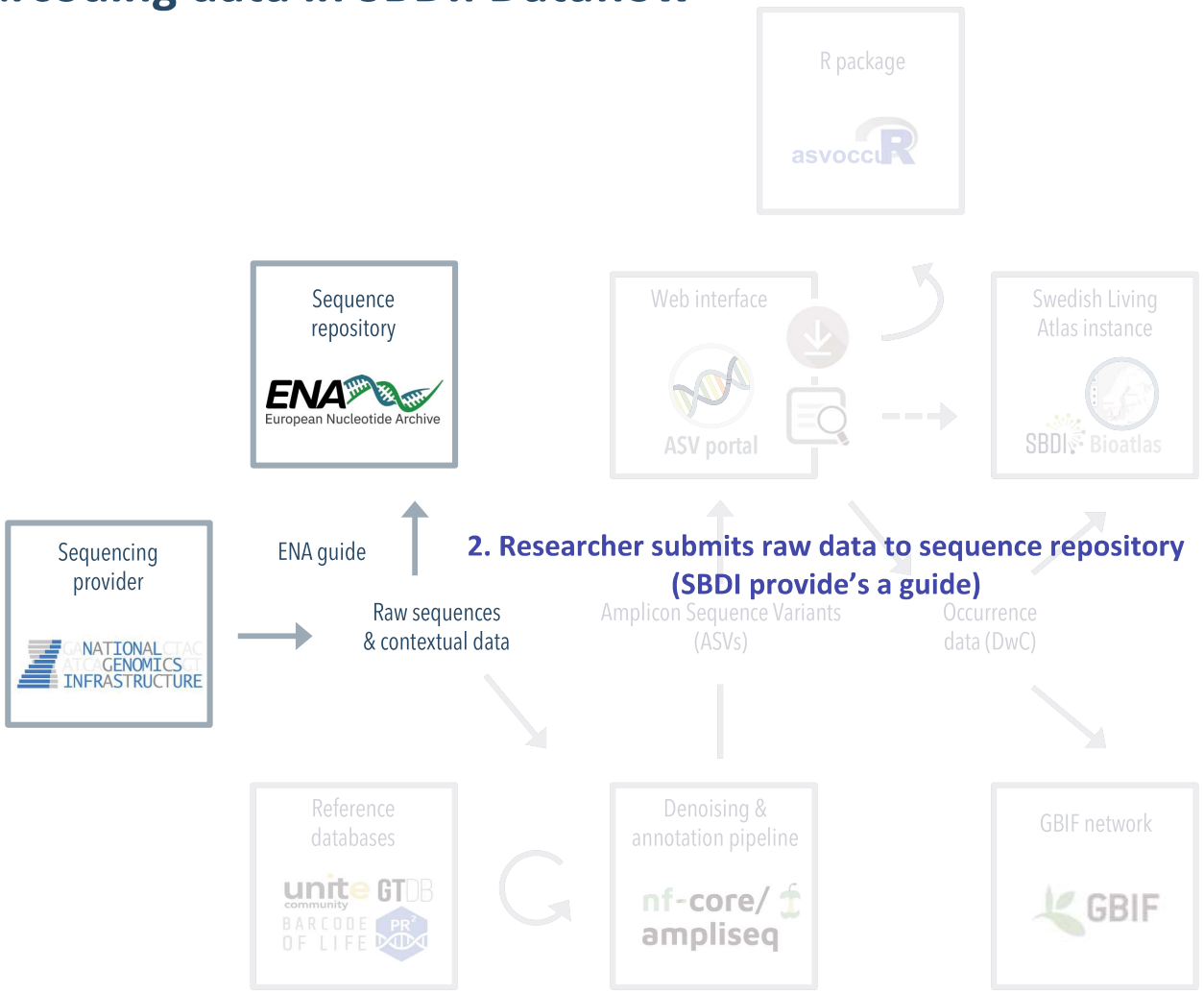
ASV counts per sample

Metabarcoding data in SBDI: Dataflow

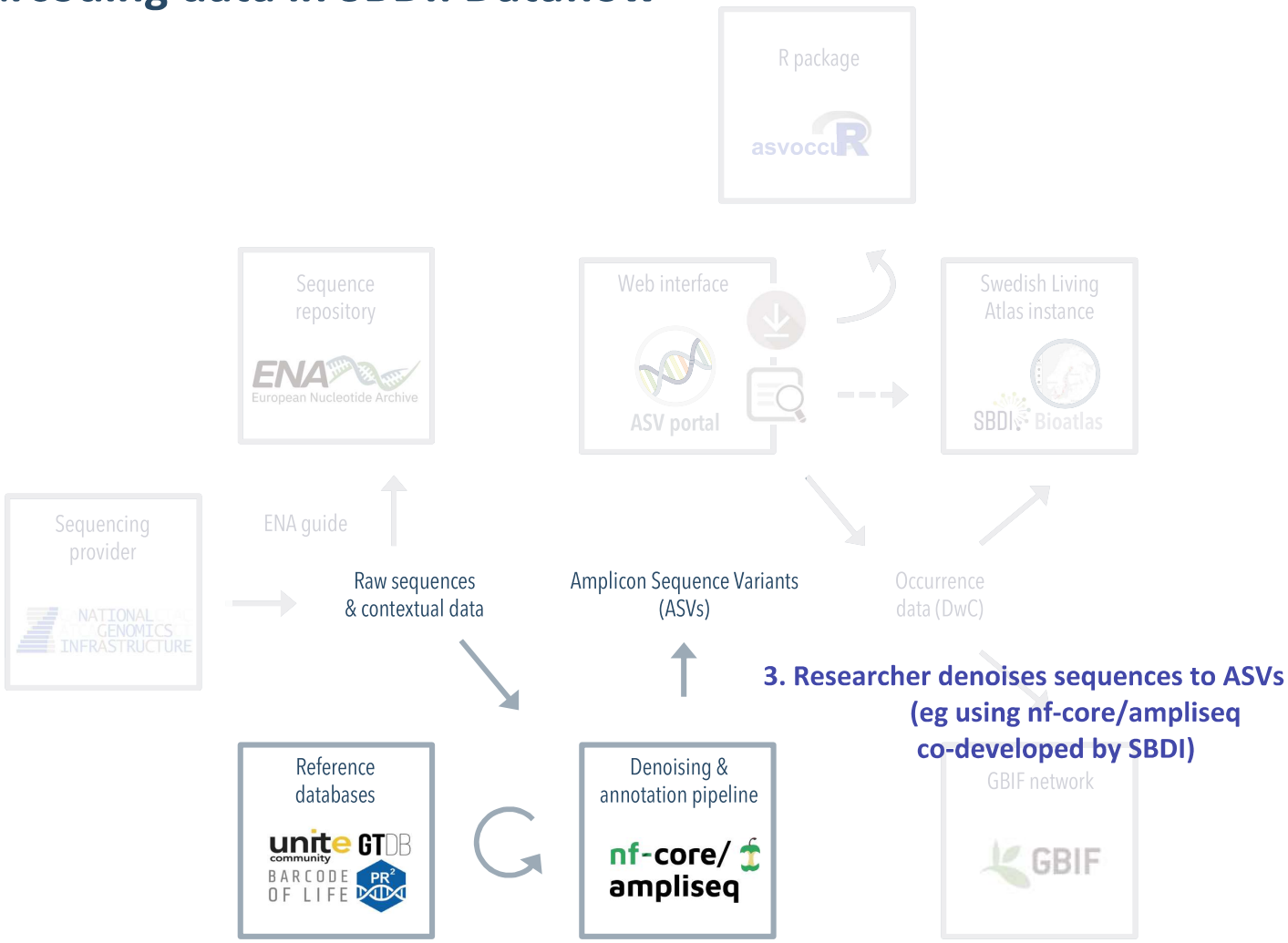
Metabarcoding data in SBDI: Dataflow



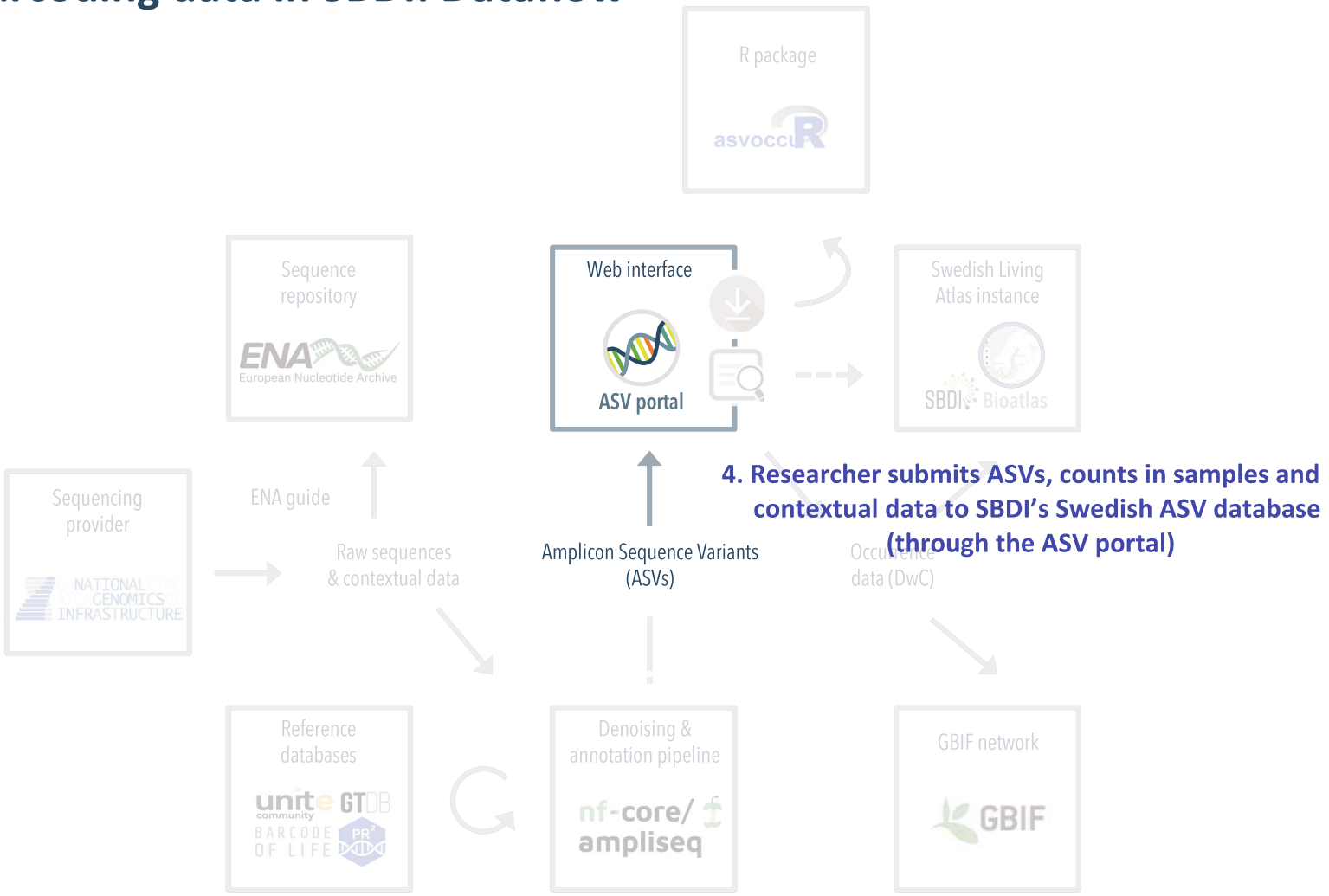
Metabarcoding data in SBDI: Dataflow



Metabarcoding data in SBDI: Dataflow

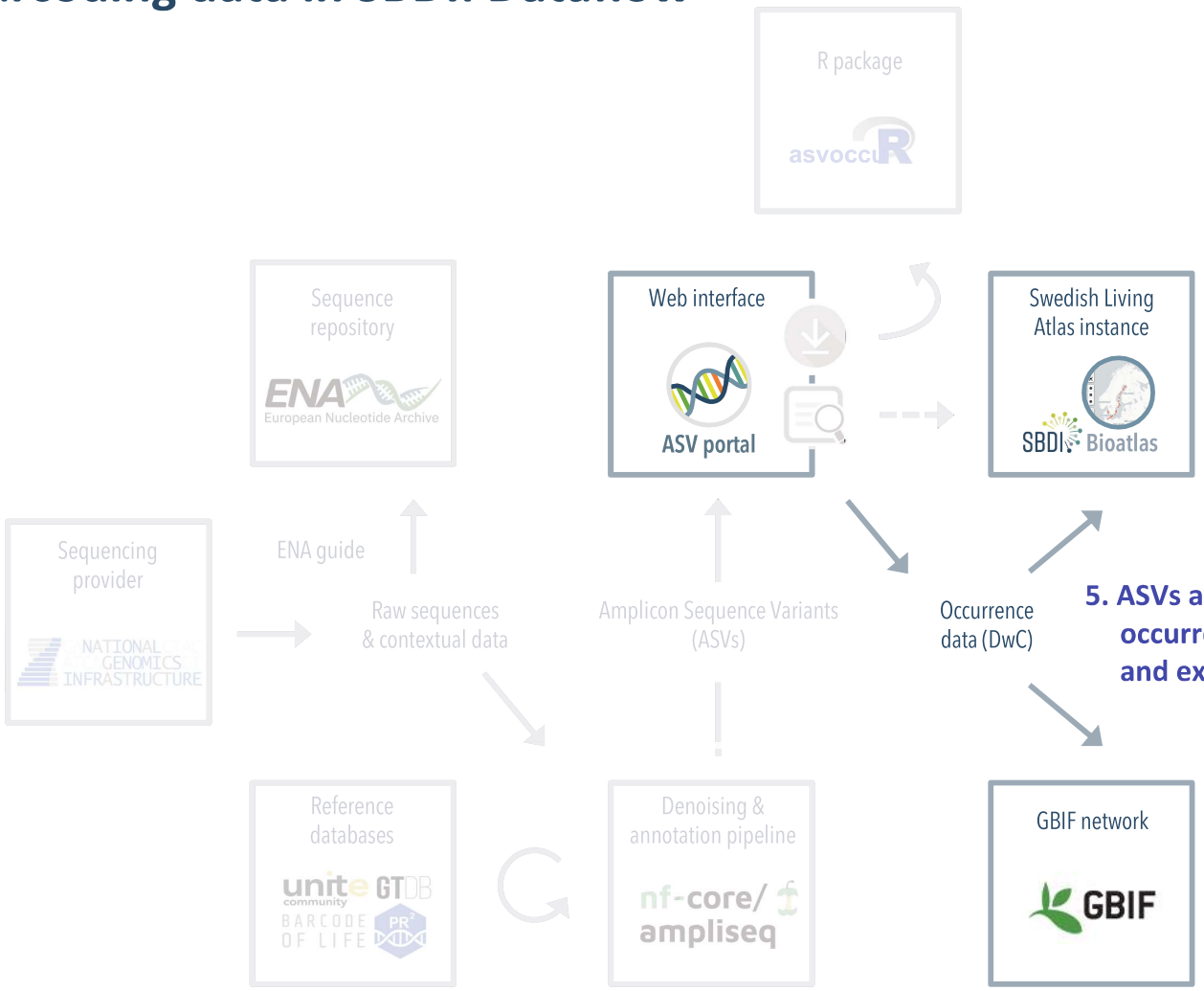


Metabarcoding data in SBDI: Dataflow



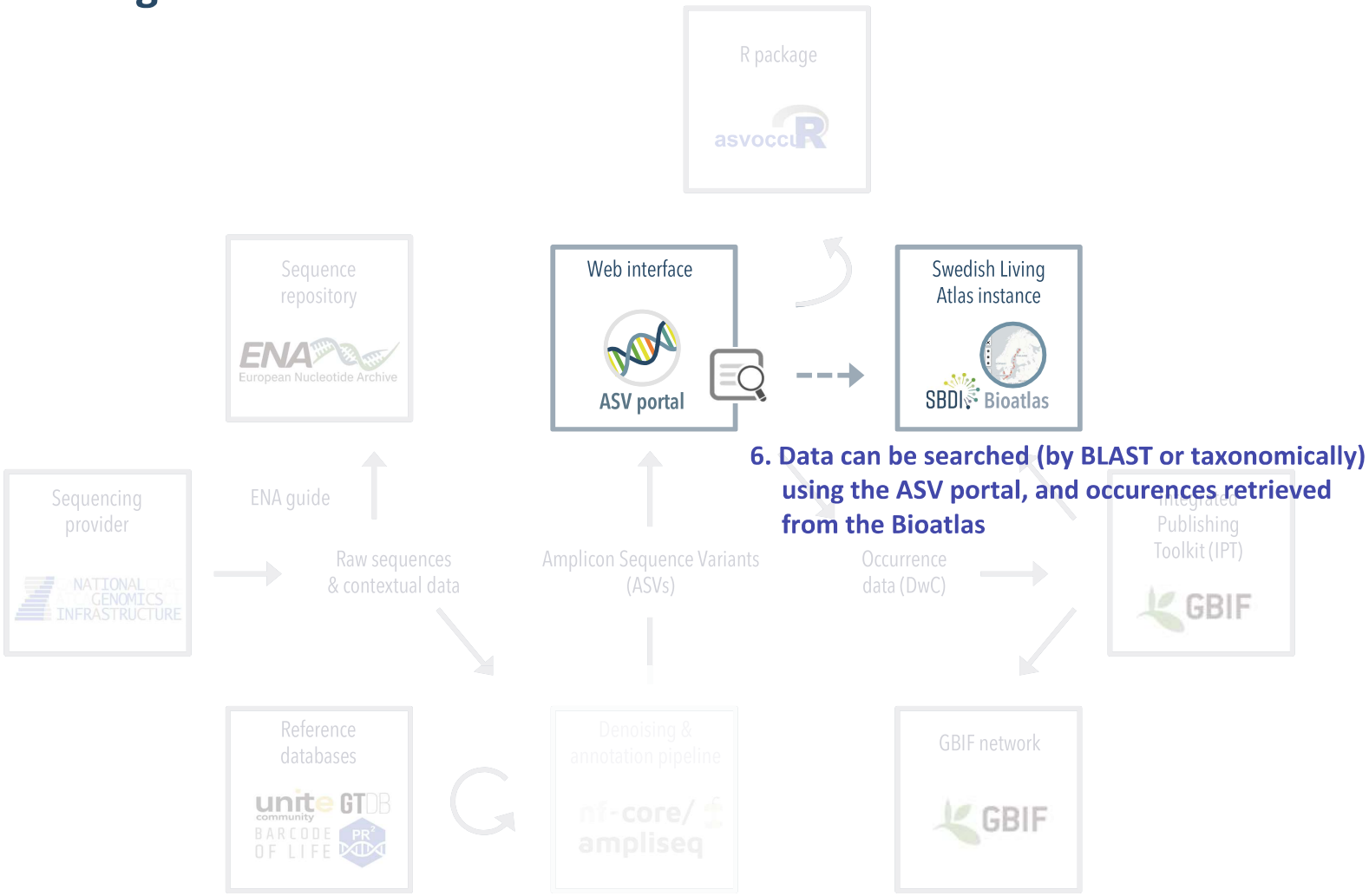
4. Researcher submits ASVs, counts in samples and contextual data to SBDI's Swedish ASV database (through the ASV portal)

Metabarcoding data in SBDI: Dataflow



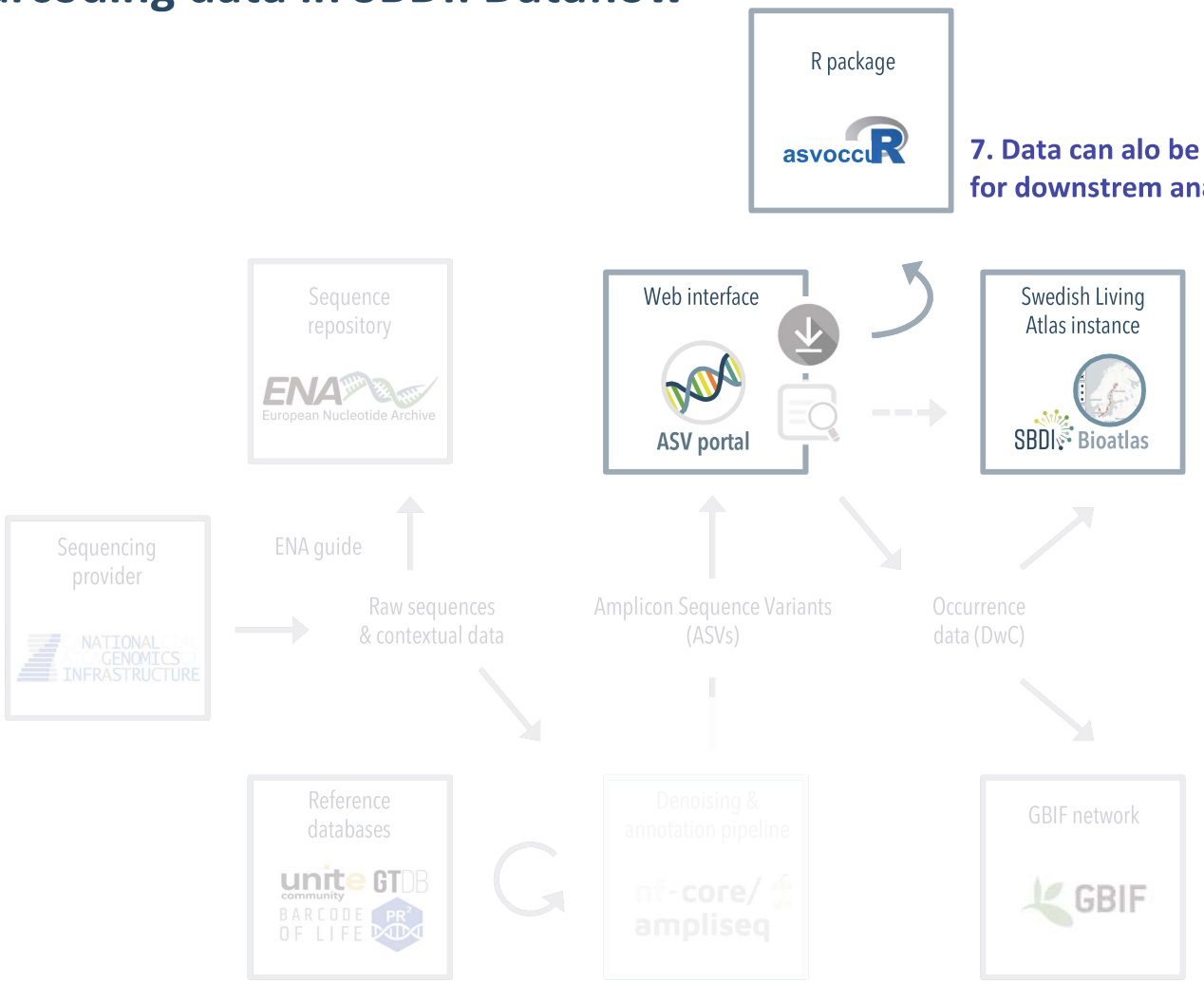
5. ASVs are taxonomically annotated and occurrence data integrated in the SBDI Bioatlas and exported to GBIF, if opted for

Metabarcoding data in SBDI: Dataflow



6. Data can be searched (by BLAST or taxonomically) using the ASV portal, and occurences retrieved from the Bioatlas

Metabarcoding data in SBDI: Dataflow

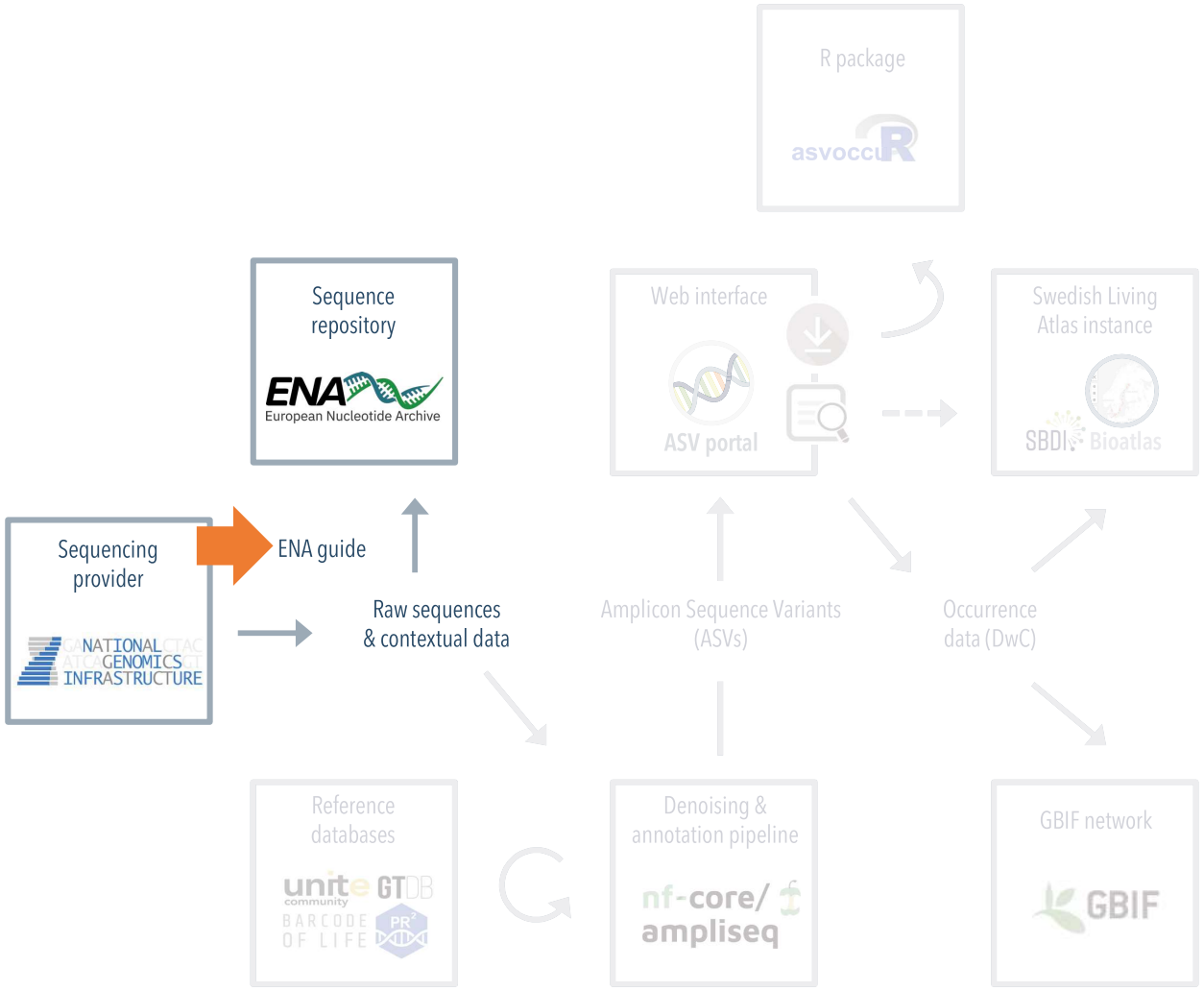


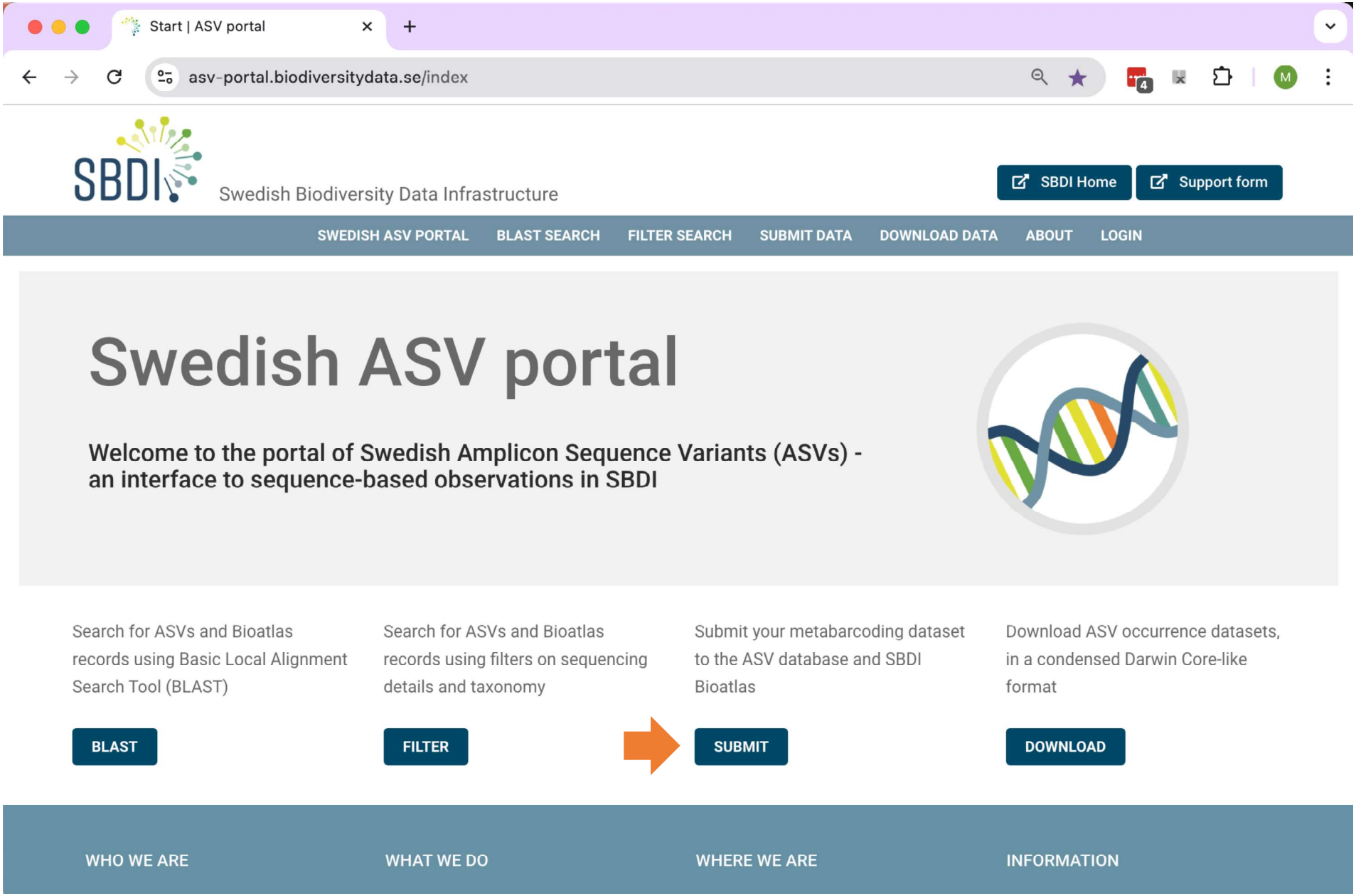
7. Data can alo be downloaded and imported in R for downstrem analysis with the asvoccur R-package

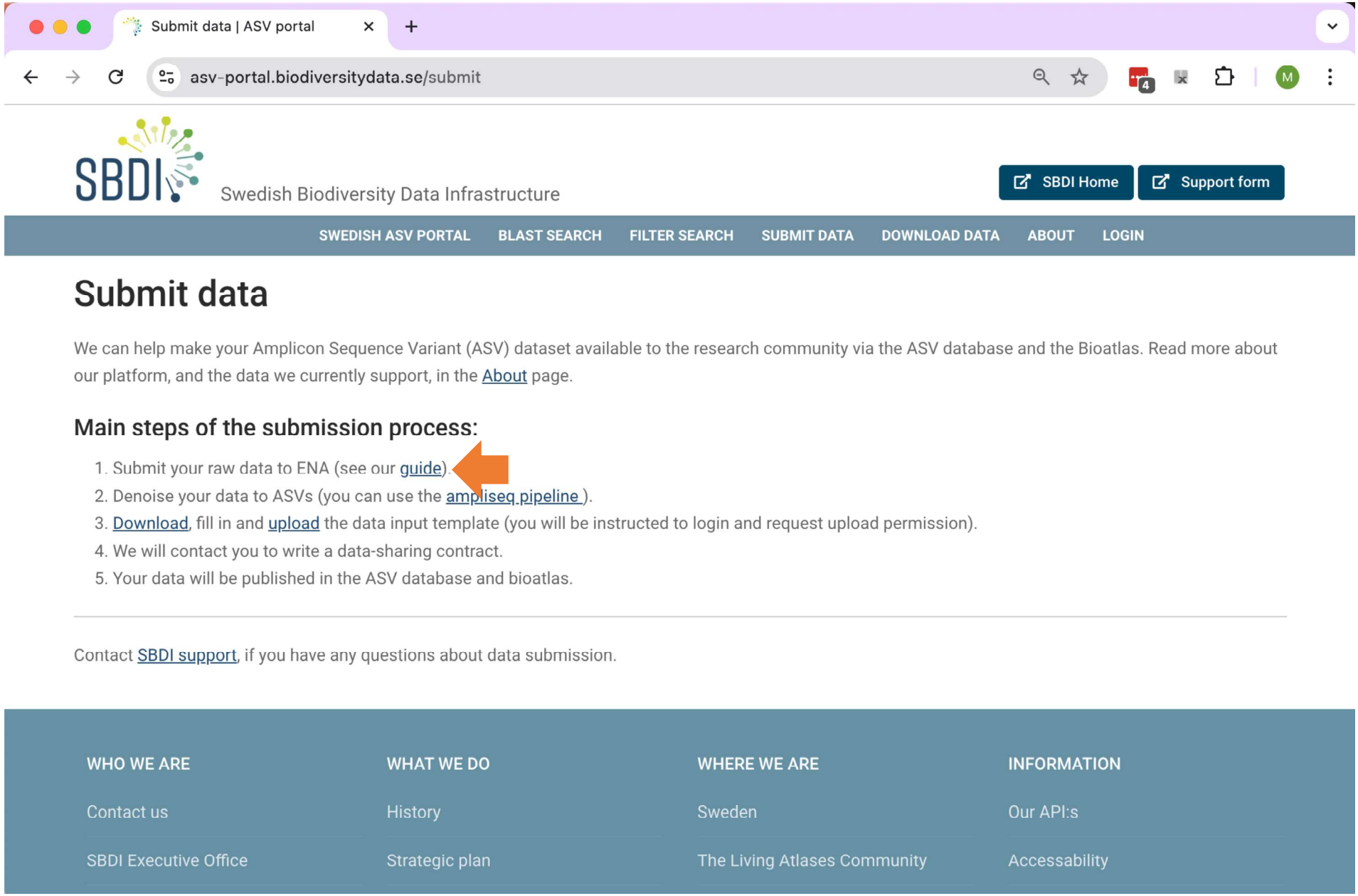
Why publish metabarcoding data in SBDI?

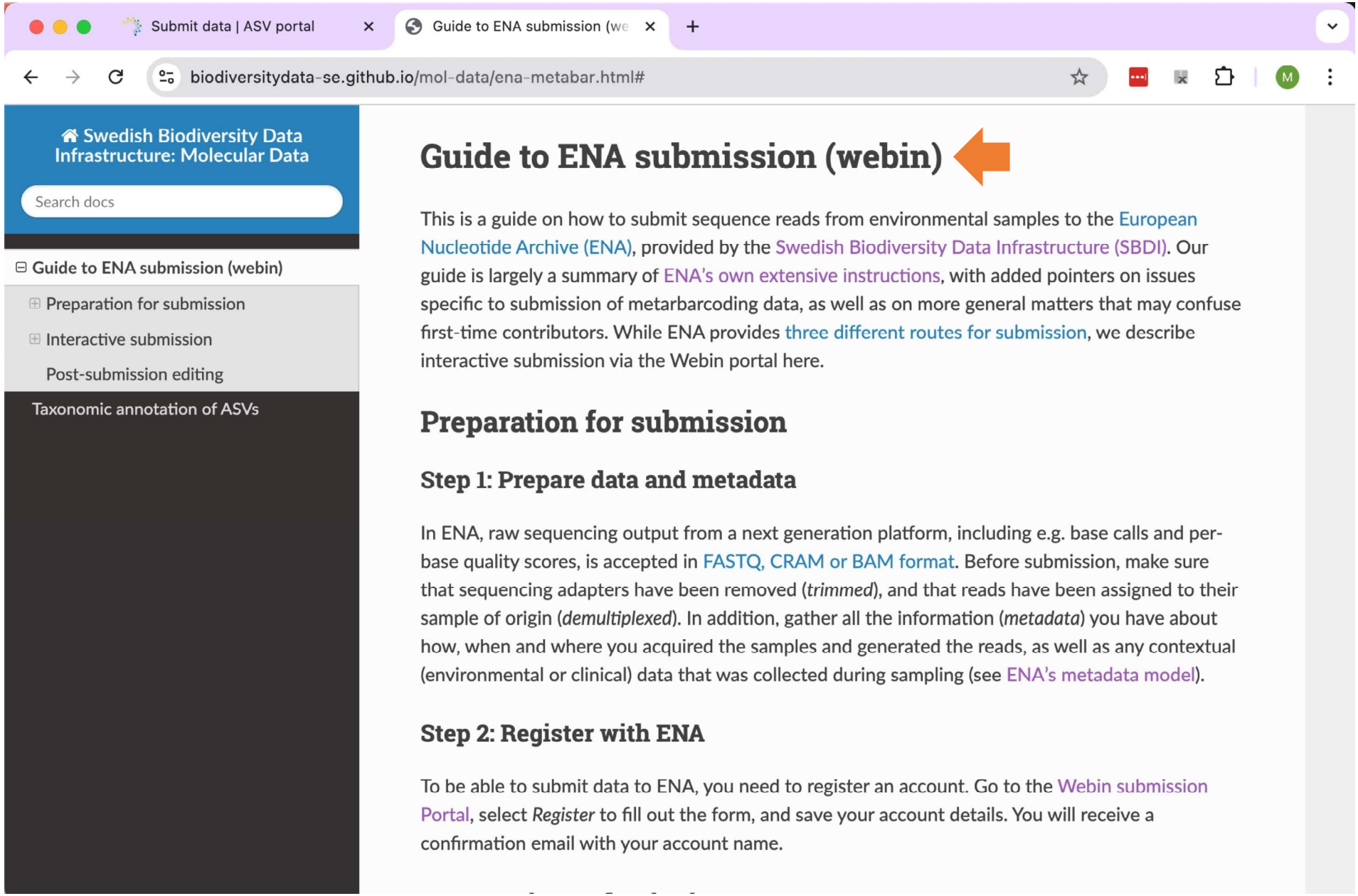
- Your data can be reused by others – you comply with the FAIR principles
- Your data can be easily compared with other metabarcoding datasets, and other types of data, in the Bioatlas
- Your submitted ASVs will be regularly reannotated by SBDI as reference databases are updated
- SBDI data is integrated in GBIF* – your data will contribute to global knowledge on biodiversity
- Your submitted data will get a DOI that can be provided in publications (complement to accession number for raw data in ENA/NCBI)

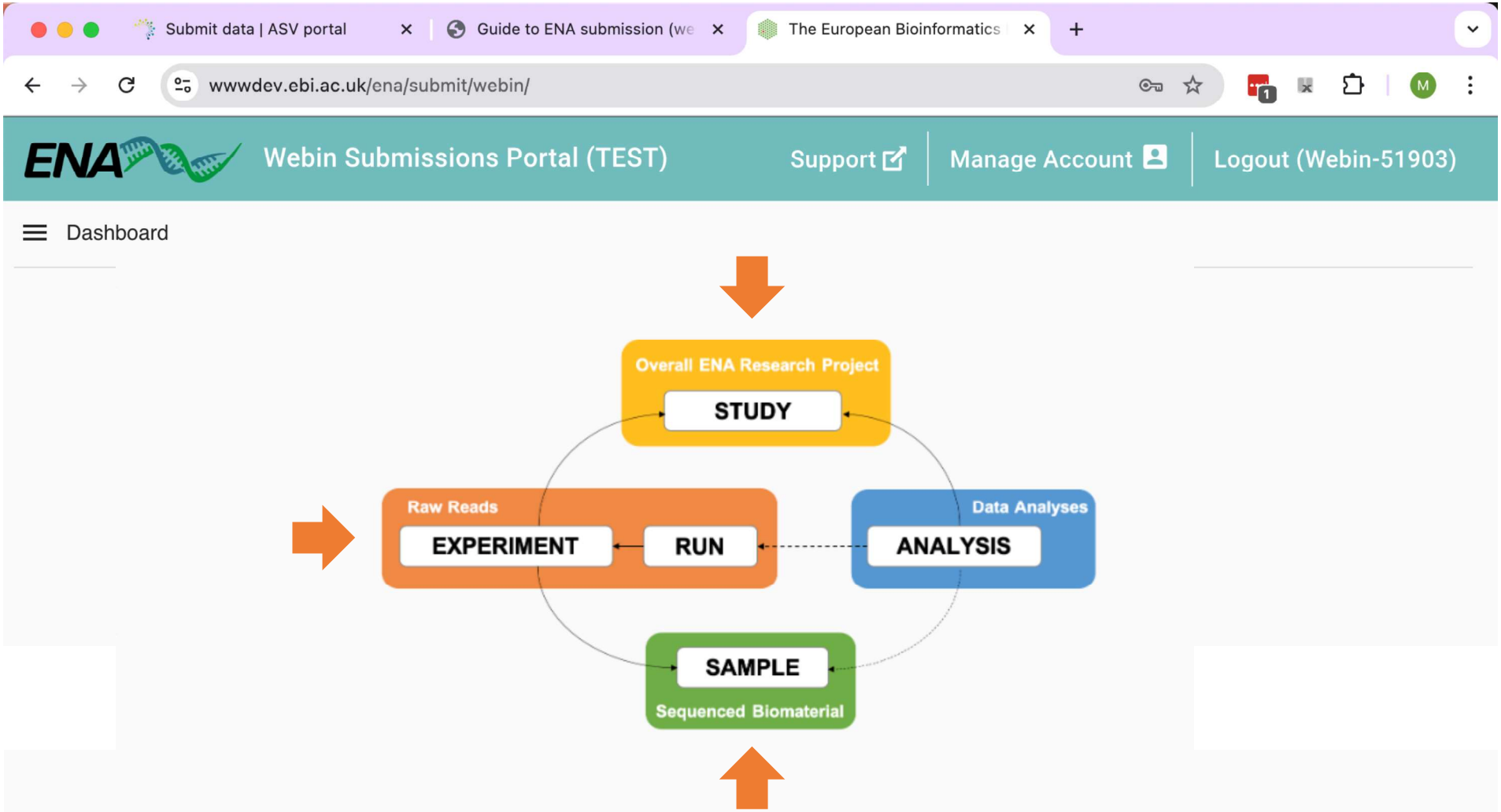
*if you opt for export to GBIF when submitting data

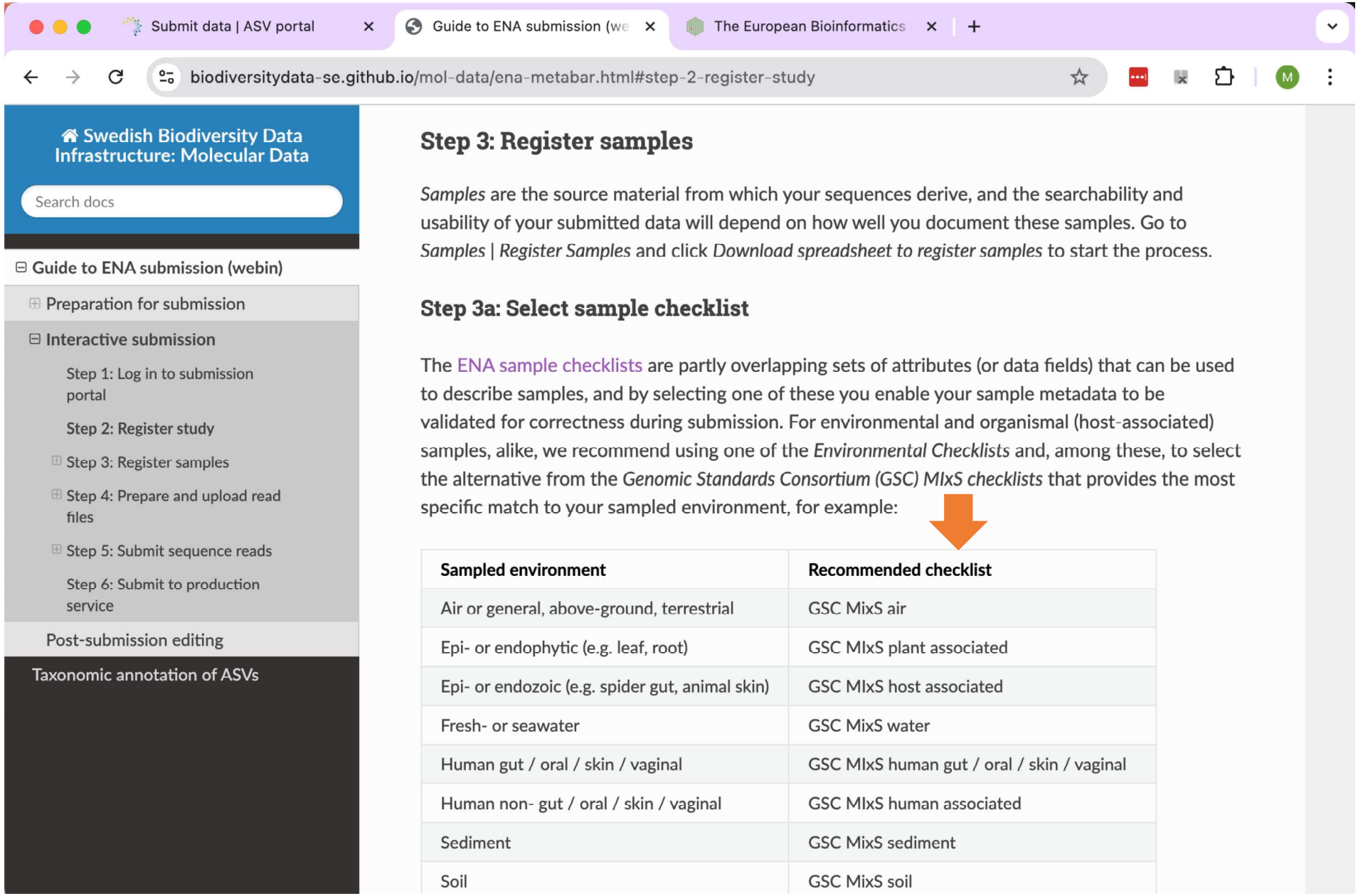

R package
asvoccuR
Sequence repository
ENA
European Nucleotide Archive
Web interface
ASV portal
Swedish Living Atlas instance
SBDI Bioatlas
Sequencing provider
NATIONAL GENOMICS INFRASTRUCTURE
ENA guide
Raw sequences & contextual data
Amplicon Sequence Variants (ASVs)
Occurrence data (DwC)
Reference databases
unite community
GTDB
BARCODE OF LIFE
Denoising & annotation pipeline
nf-core/ ampliseq
GBIF network
GBIF

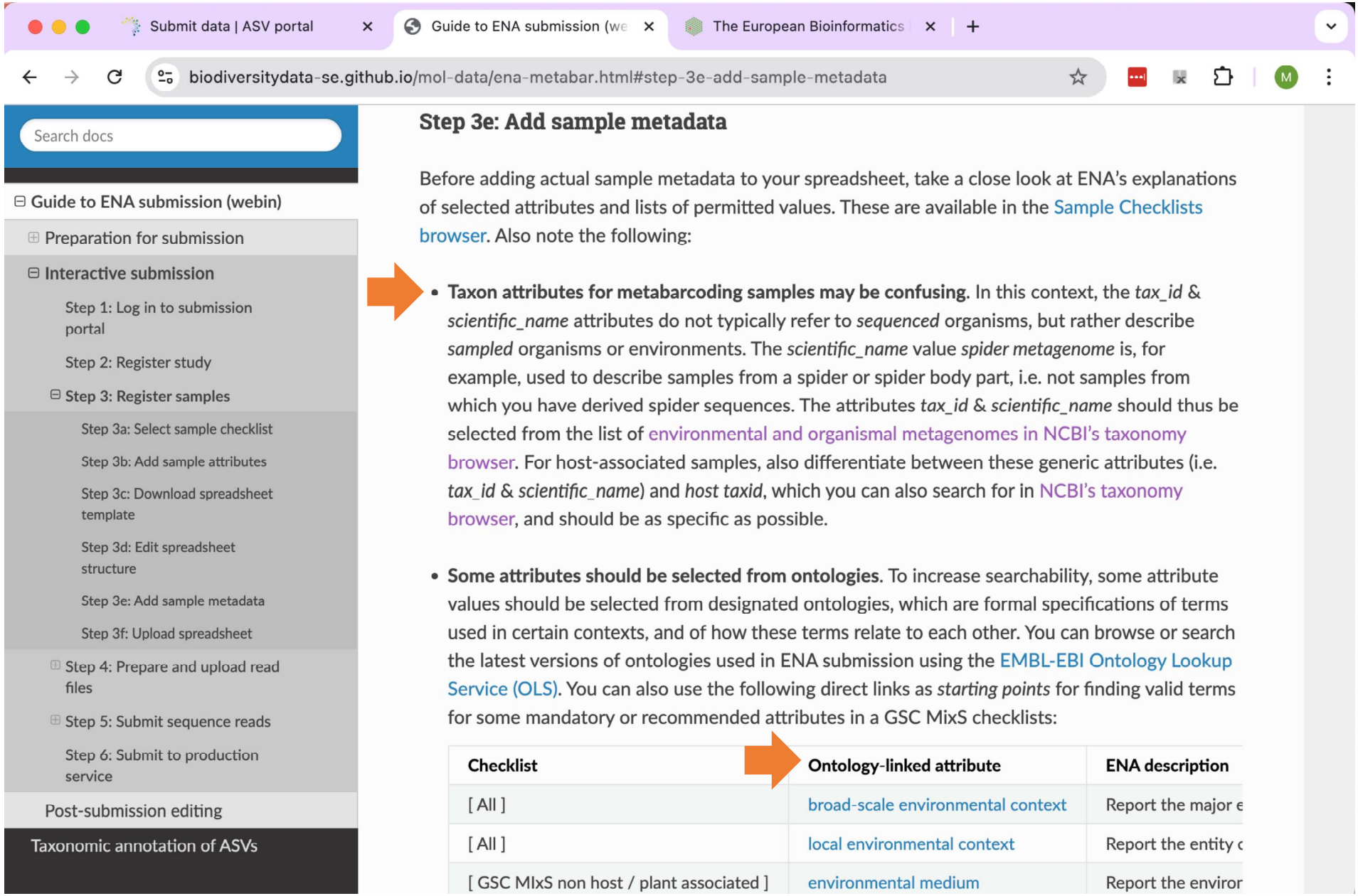

Start | ASV portal
asv-portal.biodiversitydata.se/index
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOGIN
Swedish ASV portal
Welcome to the portal of Swedish Amplicon Sequence Variants (ASVs) - an interface to sequence-based observations in SBDI
Search for ASVs and Bioatlas records using Basic Local Alignment Search Tool (BLAST)
BLAST
Search for ASVs and Bioatlas records using filters on sequencing details and taxonomy
FILTER
Submit your metabarcoding dataset to the ASV database and SBDI Bioatlas
SUBMIT
Download ASV occurrence datasets, in a condensed Darwin Core-like format
DOWNLOAD
WHO WE ARE
WHAT WE DO
WHERE WE ARE
INFORMATION

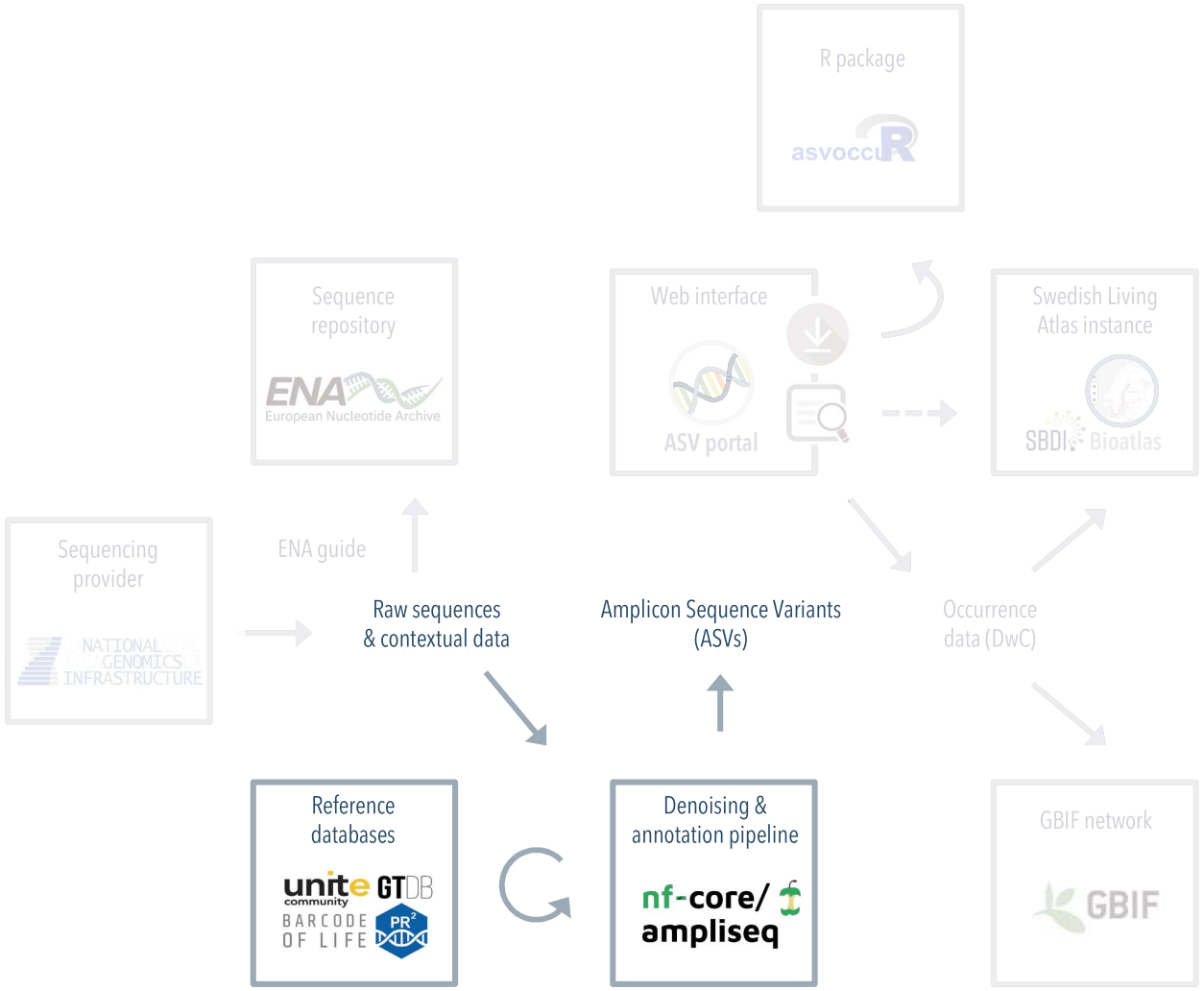

Submit data | ASV portal
asv-portal.biodiversitydata.se/submit
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOGIN
Submit data
We can help make your Amplicon Sequence Variant (ASV) dataset available to the research community via the ASV database and the Bioatlas. Read more about our platform, and the data we currently support, in the About page.
Main steps of the submission process:
1. Submit your raw data to ENA (see our guide).
2. Denoise your data to ASVs (you can use the ampliseq pipeline).
3. Download, fill in and upload the data input template (you will be instructed to login and request upload permission).
4. We will contact you to write a data-sharing contract.
5. Your data will be published in the ASV database and bioatlas.
Contact SBDI support, if you have any questions about data submission.
WHO WE ARE
Contact us
SBDI Executive Office
WHAT WE DO
History
Strategic plan
WHERE WE ARE
Sweden
The Living Atlases Community
INFORMATION
Our API:s
Accessability


Submit data | ASV portal
Guide to ENA submission (we
biodiversitydata-se.github.io/mol-data/ena-metabar.html#
Swedish Biodiversity Data Infrastructure: Molecular Data
Search docs
Guide to ENA submission (webin)
Preparation for submission
Interactive submission
Post-submission editing
Taxonomic annotation of ASVs
Guide to ENA submission (webin)
This is a guide on how to submit sequence reads from environmental samples to the European Nucleotide Archive (ENA), provided by the Swedish Biodiversity Data Infrastructure (SBDI). Our guide is largely a summary of ENA's own extensive instructions, with added pointers on issues specific to submission of metarbarcoding data, as well as on more general matters that may confuse first-time contributors. While ENA provides three different routes for submission, we describe interactive submission via the Webin portal here.
Preparation for submission
Step 1: Prepare data and metadata
In ENA, raw sequencing output from a next generation platform, including e.g. base calls and per-base quality scores, is accepted in FASTQ, CRAM or BAM format. Before submission, make sure that sequencing adapters have been removed (trimmed), and that reads have been assigned to their sample of origin (demultiplexed). In addition, gather all the information (metadata) you have about how, when and where you acquired the samples and generated the reads, as well as any contextual (environmental or clinical) data that was collected during sampling (see ENA's metadata model).
Step 2: Register with ENA
To be able to submit data to ENA, you need to register an account. Go to the Webin submission Portal, select Register to fill out the form, and save your account details. You will receive a confirmation email with your account name.

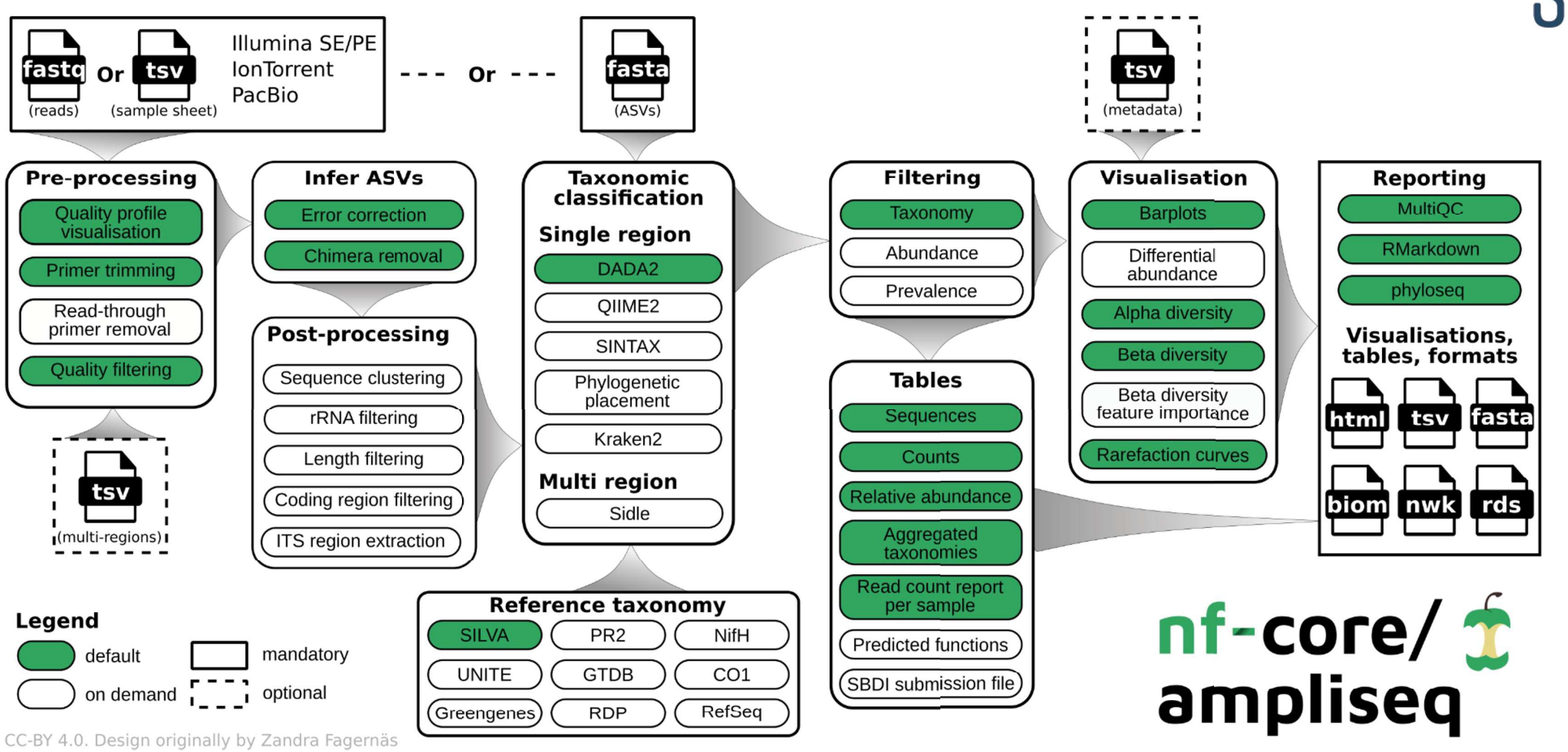

Submit data | ASV portal
Guide to ENA submission (we
The European Bioinformatics
wwwdev.ebi.ac.uk/ena/submit/webin/
ENA
Webin Submissions Portal (TEST)
Support
Manage Account
Logout (Webin-51903)
Dashboard
Overall ENA Research Project
STUDY
Raw Reads
EXPERIMENT
RUN
Data Analyses
ANALYSIS
SAMPLE
Sequenced Biomaterial


Submit data | ASV portal
Guide to ENA submission (we
The European Bioinformatics
biodiversitydata-se.github.io/mol-data/ena-metabar.html#step-2-register-study
Swedish Biodiversity Data Infrastructure: Molecular Data
Search docs
Guide to ENA submission (webin)
Preparation for submission
Interactive submission
Step 1: Log in to submission portal
Step 2: Register study
Step 3: Register samples
Step 4: Prepare and upload read files
Step 5: Submit sequence reads
Step 6: Submit to production service
Post-submission editing
Taxonomic annotation of ASVs
Step 3: Register samples
Samples are the source material from which your sequences derive, and the searchability and usability of your submitted data will depend on how well you document these samples. Go to Samples | Register Samples and click Download spreadsheet to register samples to start the process.
Step 3a: Select sample checklist
The ENA sample checklists are partly overlapping sets of attributes (or data fields) that can be used to describe samples, and by selecting one of these you enable your sample metadata to be validated for correctness during submission. For environmental and organismal (host-associated) samples, alike, we recommend using one of the Environmental Checklists and, among these, to select the alternative from the Genomic Standards Consortium (GSC) MIxS checklists that provides the most specific match to your sampled environment, for example:
Sampled environment
Recommended checklist
Air or general, above-ground, terrestrial
GSC MixS air
Epi- or endophytic (e.g. leaf, root)
GSC MIxS plant associated
Epi- or endozoic (e.g. spider gut, animal skin)
GSC MIxS host associated
Fresh- or seawater
GSC MixS water
Human gut / oral / skin / vaginal
GSC MIxS human gut / oral / skin / vaginal
Human non- gut / oral / skin / vaginal
GSC MIxS human associated
Sediment
GSC MixS sediment
Soil
GSC MixS soil


Submit data | ASV portal
Guide to ENA submission (we
The European Bioinformatics
biodiversitydata-se.github.io/mol-data/ena-metabar.html#step-3e-add-sample-metadata
Search docs
Guide to ENA submission (webin)
Preparation for submission
Interactive submission
Step 1: Log in to submission portal
Step 2: Register study
Step 3: Register samples
Step 3a: Select sample checklist
Step 3b: Add sample attributes
Step 3c: Download spreadsheet template
Step 3d: Edit spreadsheet structure
Step 3e: Add sample metadata
Step 3f: Upload spreadsheet
Step 4: Prepare and upload read files
Step 5: Submit sequence reads
Step 6: Submit to production service
Post-submission editing
Taxonomic annotation of ASVs
Step 3e: Add sample metadata
Before adding actual sample metadata to your spreadsheet, take a close look at ENA's explanations of selected attributes and lists of permitted values. These are available in the Sample Checklists browser. Also note the following:
Taxon attributes for metabarcoding samples may be confusing. In this context, the tax_id & scientific_name attributes do not typically refer to sequenced organisms, but rather describe sampled organisms or environments. The scientific_name value spider metagenome is, for example, used to describe samples from a spider or spider body part, i.e. not samples from which you have derived spider sequences. The attributes tax_id & scientific_name should thus be selected from the list of environmental and organismal metagenomes in NCBI's taxonomy browser. For host-associated samples, also differentiate between these generic attributes (i.e. tax_id & scientific_name) and host taxid, which you can also search for in NCBI's taxonomy browser, and should be as specific as possible.
Some attributes should be selected from ontologies. To increase searchability, some attribute values should be selected from designated ontologies, which are formal specifications of terms used in certain contexts, and of how these terms relate to each other. You can browse or search the latest versions of ontologies used in ENA submission using the EMBL-EBI Ontology Lookup Service (OLS). You can also use the following direct links as starting points for finding valid terms for some mandatory or recommended attributes in a GSC MixS checklists:
Checklist
Ontology-linked attribute
ENA description
[All]
broad-scale environmental context
Report the major e
[All]
local environmental context
Report the entity
[GSC MIxS non host / plant associated]
environmental medium
Report the enviror

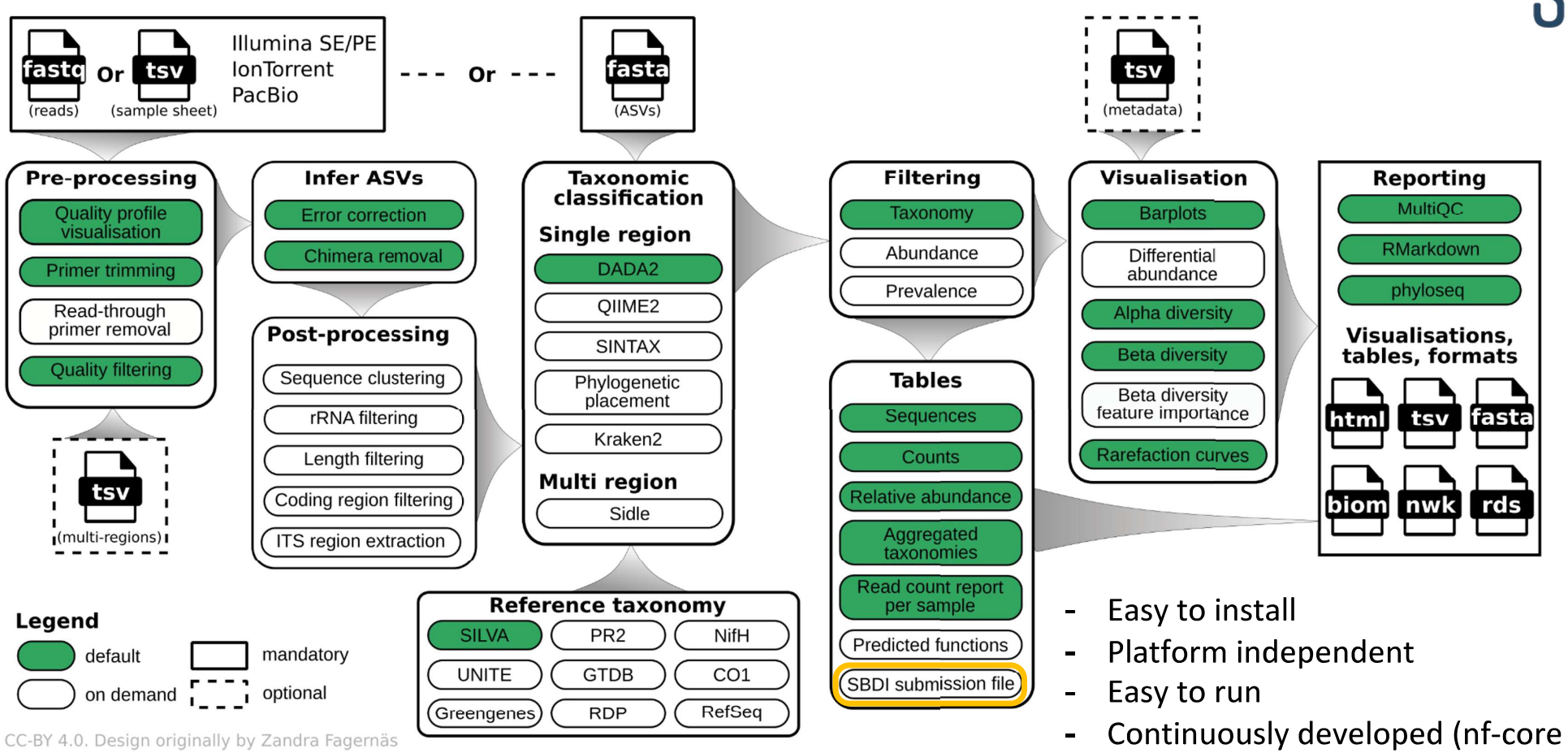

R package
asvoccuR
Sequence repository
ENA
European Nucleotide Archive
Web interface
ASV portal
Swedish Living Atlas instance
SBDI Bioatlas
Sequencing provider
NATIONAL GENOMICS INFRASTRUCTURE
ENA guide
Raw sequences & contextual data
Amplicon Sequence Variants (ASVs)
Occurrence data (DwC)
Reference databases
unite community
GTDB
BARCODE OF LIFE
PR²
Denoising & annotation pipeline
nf-core/ ampliseq
GBIF network
GBIF

nf-core/ampliseq



Straub et al. Front Micr 2020

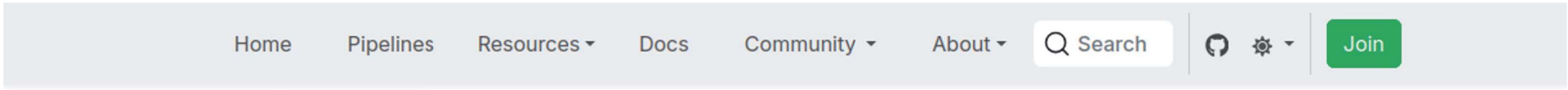
nf-core/ampliseq



- Easy to install
- Platform independent
- Easy to run
- Continuously developed (nf-core collaboration)

Straub et al. Front Micr 2020

nf-core/ampliseq: running through command line



Running the pipeline

Quick start

The typical command for running the pipeline is as follows:

```
nextflow run nf-core/ampliseq \
    -r 2.7.1 \
    -profile singularity \
    --input "samplesheet.tsv" \
    --FW_primer GTGYCAGCMGCCGCGGTAA \
    --RV_primer GGACTACNVGGGTWTCTAAT \
    --metadata "data/Metadata.tsv" \
    --outdir "./results"
```

In this example, `--input` is the Samplesheet input, other options are Direct FASTQ input and ASV/OTU fasta input. For more details on metadata, see Metadata. For Reproducibility, specify the version to run using `-r` (= release, e.g. 2.7.1, please use the most recent release). See the nf-core/ampliseq website documentation for more information about pipeline specific parameters.

It is possible to not provide primer sequences (`--FW_primer` & `--RV_primer`) and skip primer trimming using `--skip_cutadapt` , but this is only for data that indeed does not contain any PCR primers in their sequences. Also, metadata (`--metadata`) isnt required, but aids downstream analysis.

On this page

Table of Contents
Running the pipeline
Core Nextflow arguments
Custom configuration
Azure Resource Requests
Running in the background
Nextflow memory requirements
Back to top
Show details

nf-core/ampliseq: What do you need?

Prerequisities (available as modules at Dardel / PDC)

- Nextflow
- Singularity/apptainer or docker (or conda)
- If you plan to run the workflow offline, you need to first download the workflow and all singularity images (or similar)

nf-core/ampliseq: What do you need?

Preparations

- Locate your raw data, recommended to create a samplesheet
- Primer sequences?
- Which parameters do you want to use?
- Optional: Collect metadata in a file
  - — sample collection information: date, location
  - — sequencing information
  - — any other data

nf-core/ampliseq: samplesheet


```
--input 'path/to/samplesheet.tsv'
```

For example, the tab-separated samplesheet may contain:

sampleID	forwardReads	reverseReads	run
sample1	./data/S1_R1_001.fastq.gz	./data/S1_R2_001.fastq.gz	A
sample2	./data/S2_fw.fastq.gz	./data/S2_rv.fastq.gz	A
sample3	./S4x.fastq.gz	./S4y.fastq.gz	B
sample4	./a.fastq.gz	./b.fastq.gz	B

Please note the following requirements:

- 2 to 4 columns/entries
- File extensions `.tsv`, `.csv`, `.yml`, `.yaml` specify the file type, otherwise file type will be derived from content, if possible
- Must contain the header `sampleID` and `forwardReads`
- May contain the header `reverseReads` and `run`
- Sample IDs must be unique
- Sample IDs must start with a letter
- Sample IDs can only contain letters, numbers or underscores
- FastQ files must be compressed (`.fastq.gz`, `.fq.gz`)
- Within one samplesheet, only one type of raw data should be specified (same amplicon & sequencing method)

An example samplesheet has been provided with the pipeline.

To avoid producing a sample sheet, Direct FASTQ input may be used instead.

On this page

Table of Contents
Running the pipeline
Core Nextflow arguments
Custom configuration
Azure Resource Requests
Running in the background
Nextflow memory requirements
Back to top

Show details

nf-core/ampliseq: setting parameters in a file

- Parameters can be collected in a yaml or json file
  - reproducible
  - easy to run again with the same settings

Command:
nextflow run nf-core/ampliseq -r 2.11.0 -profile singularity -params-file params.yaml

… where params.yaml contains:

```
input:        samplesheet.tsv
FW_primer: "GTGYCAGCMGCCGCGGTAA"
RV_primer: "GGACTACNVGGGTWTCTAAT"
metadata: "data/Metadata.tsv"
outdir: "./results"
sbdi-export: true
```

nf-core/ampliseq: setting parameters in a file

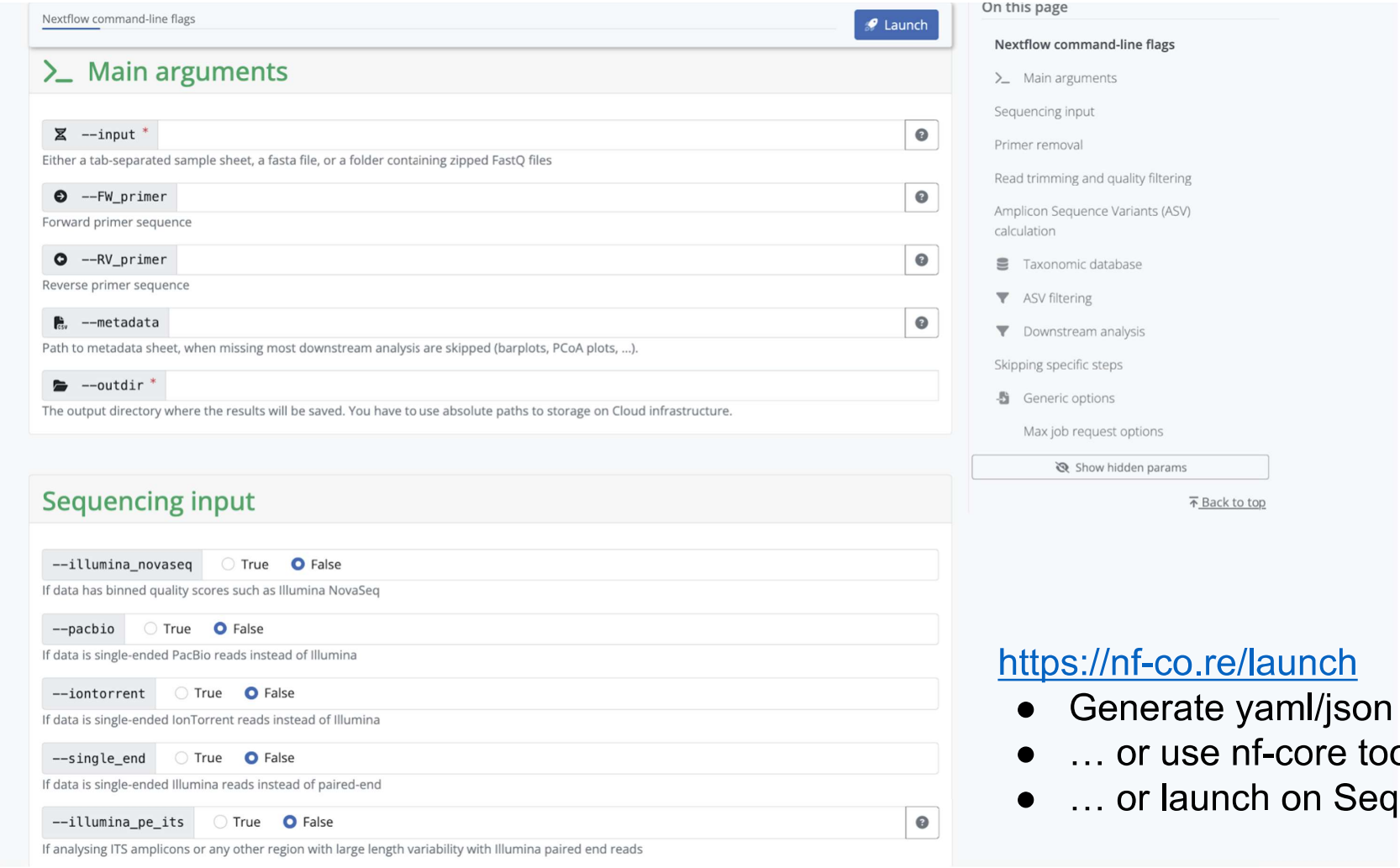
- Parameters can be collected in a yaml or json file
  - — reproducible
  - — easy to run again with the same settings

Command, if running on Dardel:
nextflow run nf-core/ampliseq -r 2.11.0 -profile **pdc_kth** -params-file params.yaml

… where params.yaml contains:

```
input:        samplesheet.tsv
FW_primer: "GTGYCAGCMGCCGCGGTAA"
RV_primer: "GGACTACNVGGGTWTCTAAT"
metadata: "data/Metadata.tsv"
outdir: "./results"
sbdi-export: true
```

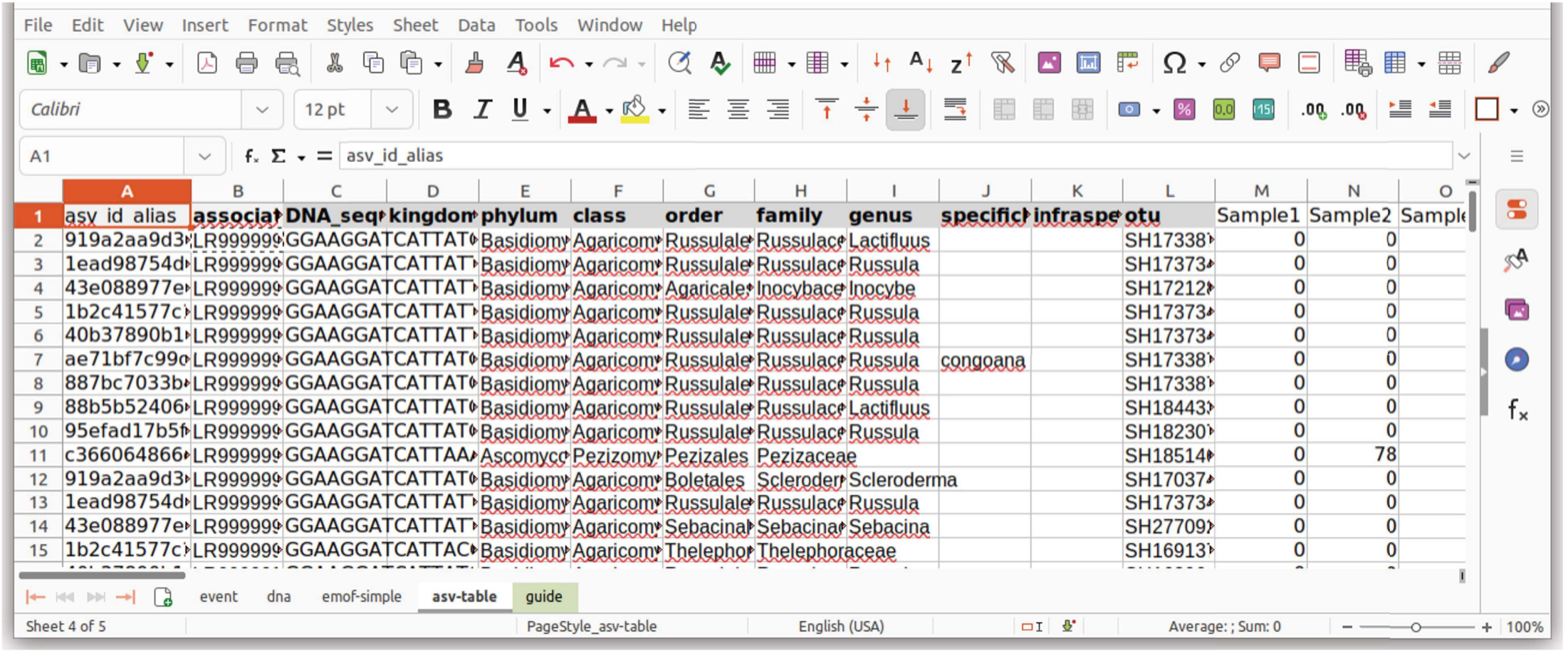
nf-core/ampliseq: running through graphical user interface



https://nf-co.re/launch

- Generate yaml/json parameters file
- … or use nf-core tools to launch workflow
- … or launch on Seqera Platform

nf-core/ampliseq: output example using sbdi-export option



File Edit View Insert Format Styles Sheet Data Tools Window Help

Calibri 12 pt

A1 asv_id_alias

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	asv_id_alias	associat	DNA_seq	kingdom	phylum	class	order	family	genus	specific	infraspe	otu	Sample1	Sample2	Sample
2	919a2aa9d3	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Lactifluus			SH17338	0	0	
3	1ead98754d	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Russula			SH17373	0	0	
4	43e088977e	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Agaricale	Inocybace	Inocybe			SH17212	0	0	
5	1b2c41577c	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Russula			SH17373	0	0	
6	40b37890b1	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Russula			SH17373	0	0	
7	ae71bf7c99	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Russula	congoana		SH17338	0	0	
8	887bc7033b	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Russula			SH17338	0	0	
9	88b5b52406	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Lactifluus			SH18443	0	0	
10	95efad17b5	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Russula			SH18230	0	0	
11	c366064866	LR999999	GGAAGGAT	CATTAA	Ascomyco	Pezizomy	Pezizales	Pezizaceae				SH18514	0	78	
12	919a2aa9d3	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Boletales	Scleroder	Scleroderma			SH17037	0	0	
13	1ead98754d	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Russulale	Russulace	Russula			SH17373	0	0	
14	43e088977e	LR999999	GGAAGGAT	CATTAT	Basidiomy	Agaricom	Sebacinal	Sebacinac	Sebacina			SH27709	0	0	
15	1b2c41577c	LR999999	GGAAGGAT	CATTAC	Basidiomy	Agaricom	Thelephor	Thelephoraceae				SH16913	0	0	

event dna emof-simple asv-table guide

Sheet 4 of 5 PageStyle_asv-table English (USA) Average: ; Sum: 0 100%

Metabarcoding data in SBDI: Supported data types

Sequencing platforms compatible with the nf-core/ampliseq pipeline

- Illumina SE/PE
- PacBio
- IonTorrent

Taxonomic groups/marker genes currently used in SBDI*

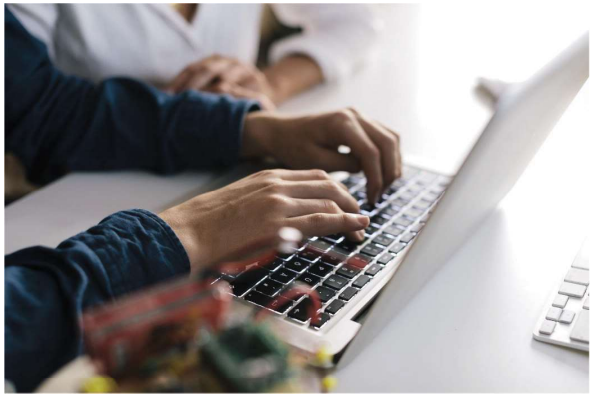
Taxa	Marker	Database
Archaea and Bacteria	16S rRNA	GTDB
Fungi	ITS	UNITE
Eukaryotes	18S rRNA	PR2
Metazoa	COI	BOLD

*Other groups/markers can be supported on request

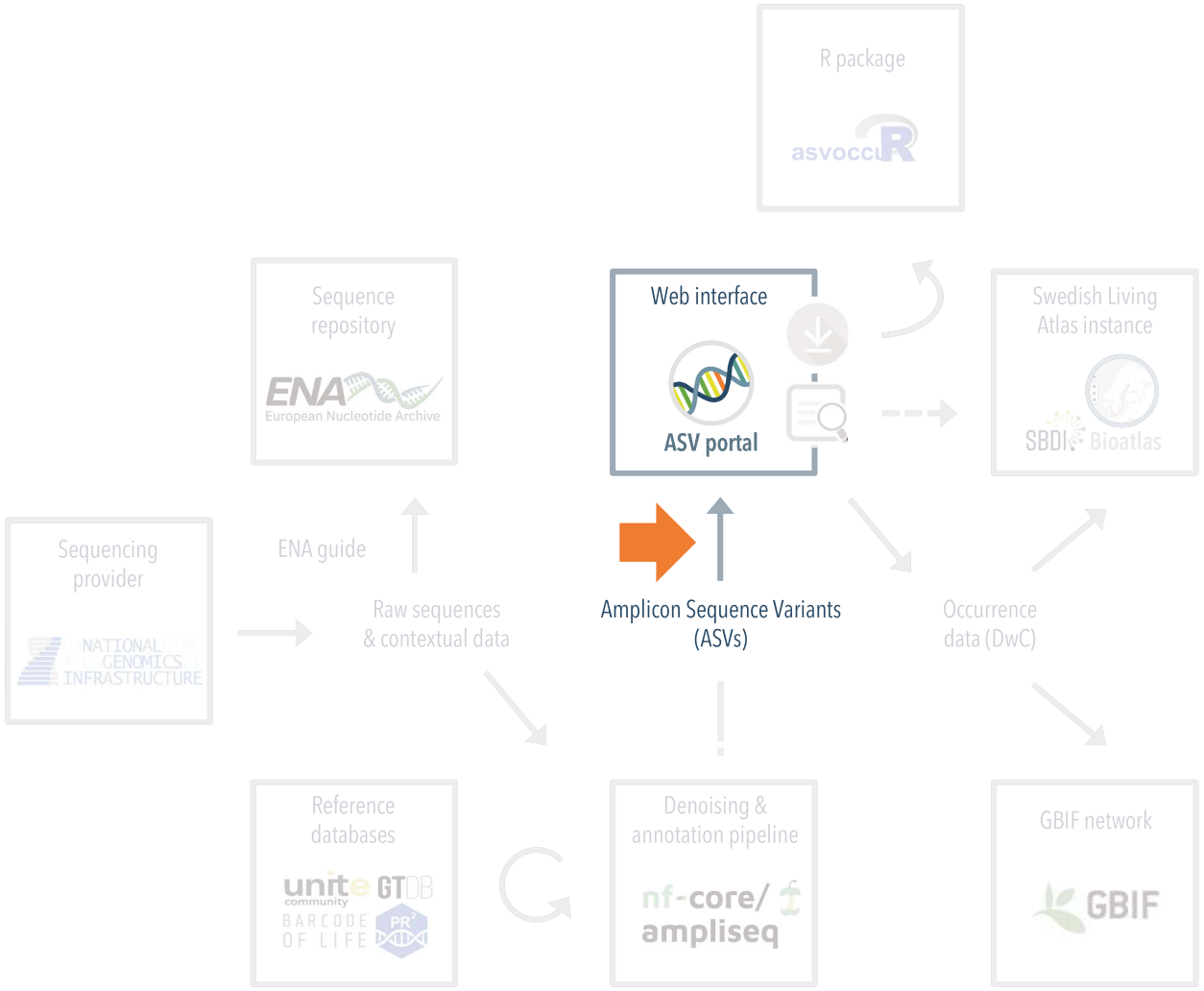
nf-core/ampliseq: getting help

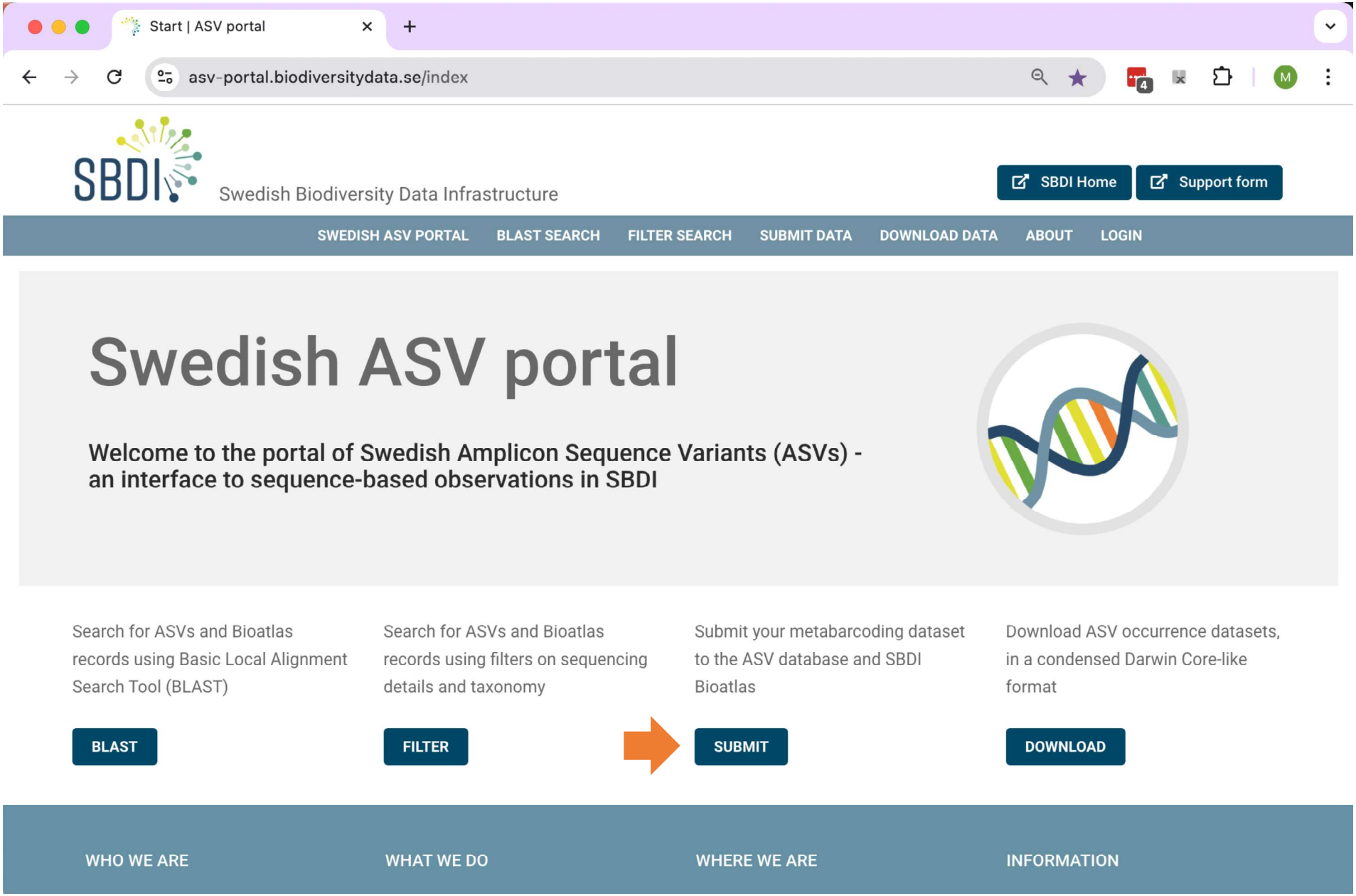
- NBIS Drop-in support
  http://meet.nbis.se/dropin

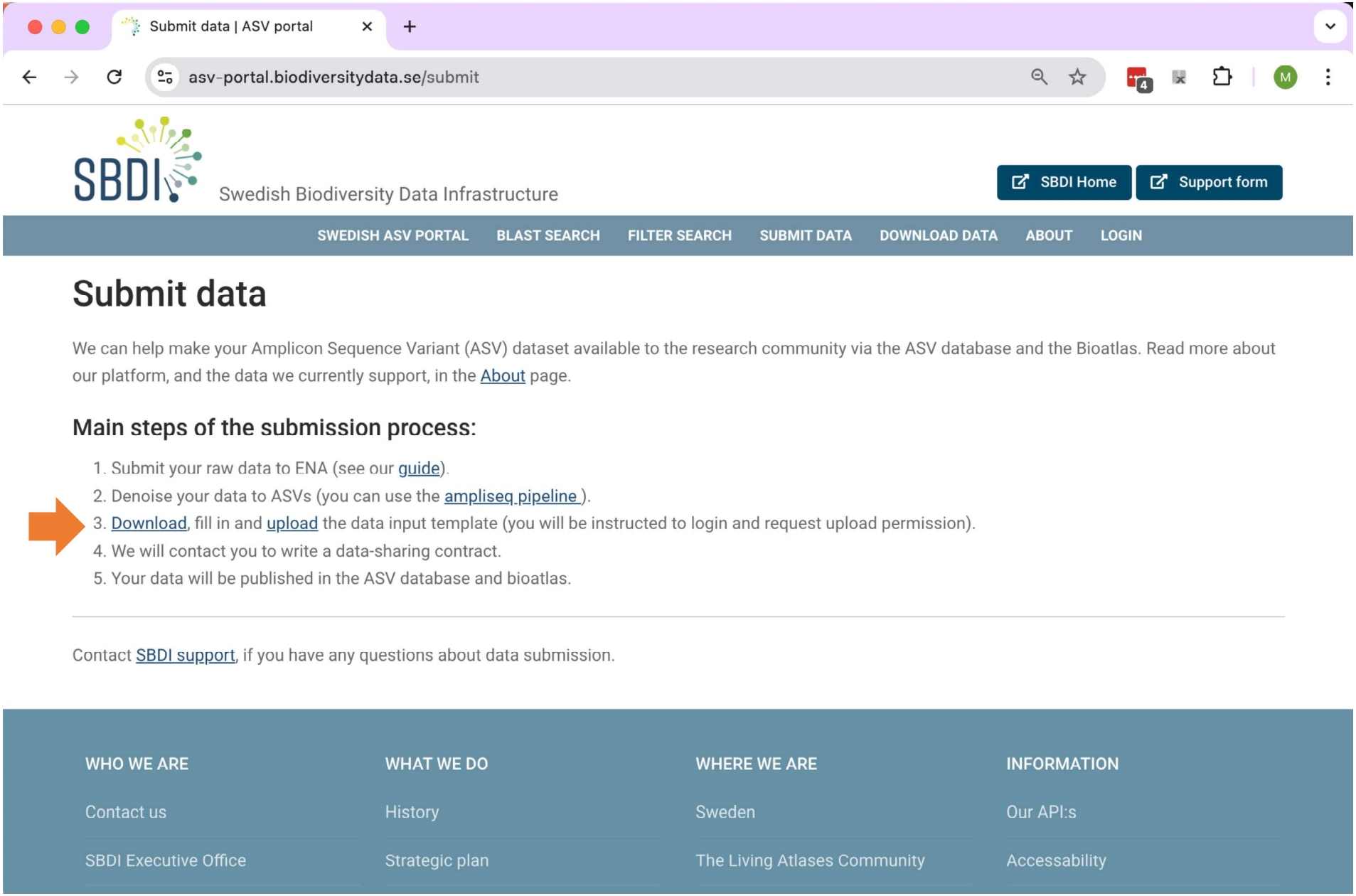
Drop-in sessions	Ask your question! Bioinformatics, sequencing, data management, and more Online **- Tuesdays at 14.00** On site - Stockholm and Lund

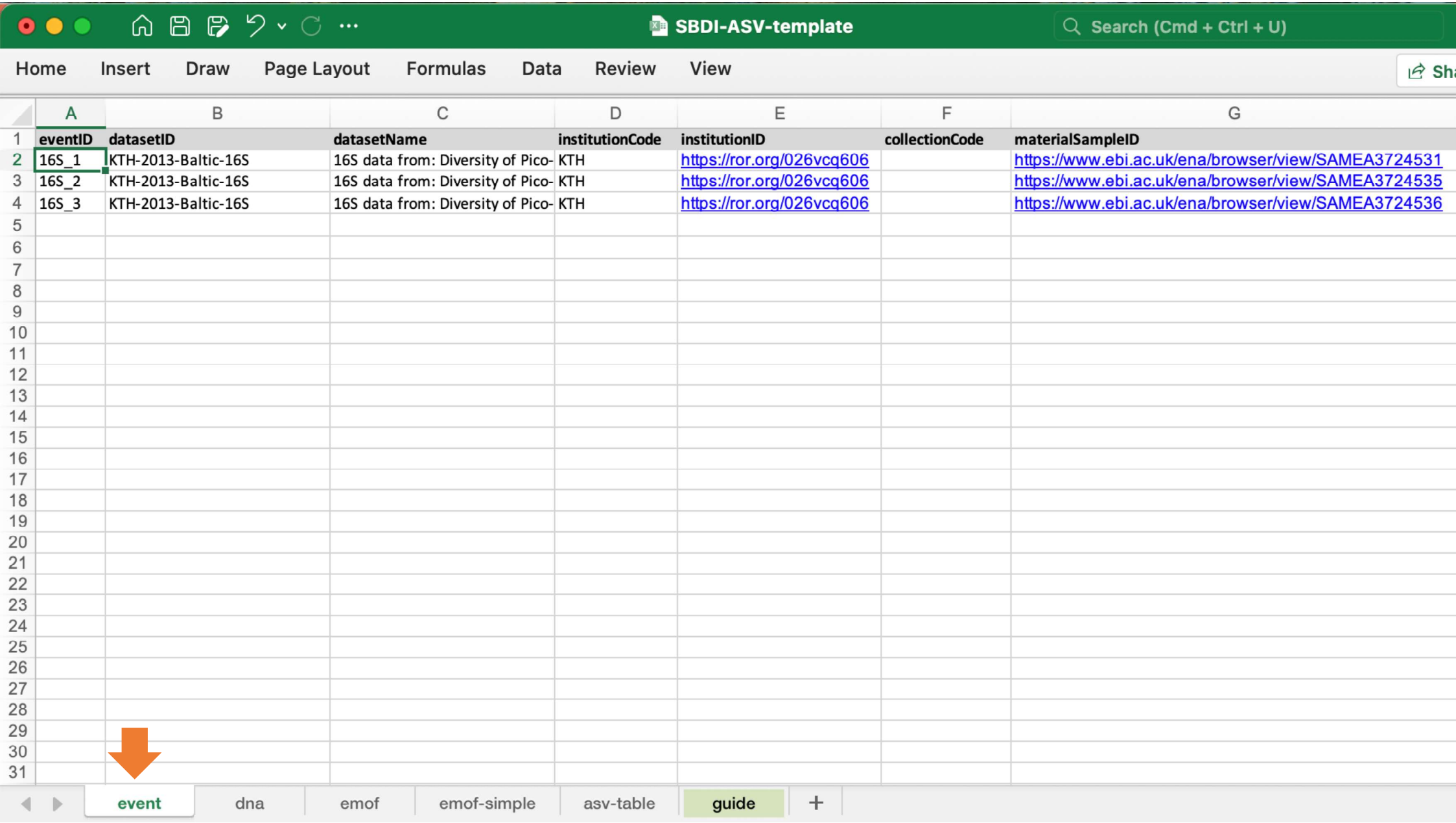


- nf-core ampliseq slack channel
  https://nfcore.slack.com/channels/ampliseq

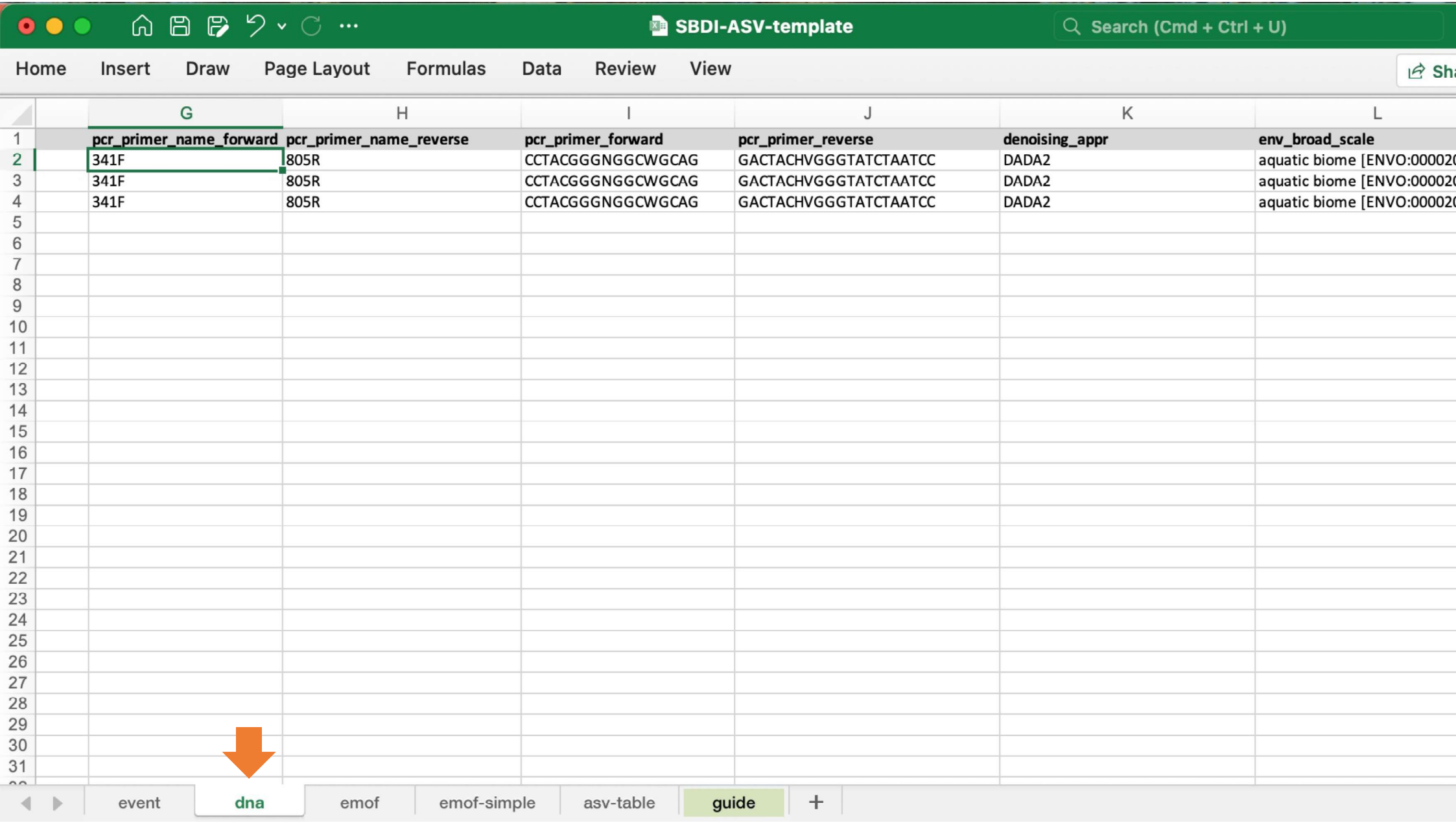

R package
asvoccuR
Sequence repository
ENA
European Nucleotide Archive
Web interface
ASV portal
Swedish Living Atlas instance
SBDI Bioatlas
Sequencing provider
NATIONAL GENOMICS INFRASTRUCTURE
ENA guide
Raw sequences & contextual data
Amplicon Sequence Variants (ASVs)
Occurrence data (DwC)
Reference databases
unite community
GTDB
BARCODE OF LIFE
Denoising & annotation pipeline
nf-core/ampliseq
GBIF network
GBIF


Start | ASV portal
asv-portal.biodiversitydata.se/index
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOGIN
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)
BLAST
Search for ASVs and Bioatlas records using filters on sequencing details and taxonomy
FILTER
Submit your metabarcoding dataset to the ASV database and SBDI Bioatlas
SUBMIT
Download ASV occurrence datasets, in a condensed Darwin Core-like format
DOWNLOAD
WHO WE ARE
WHAT WE DO
WHERE WE ARE
INFORMATION

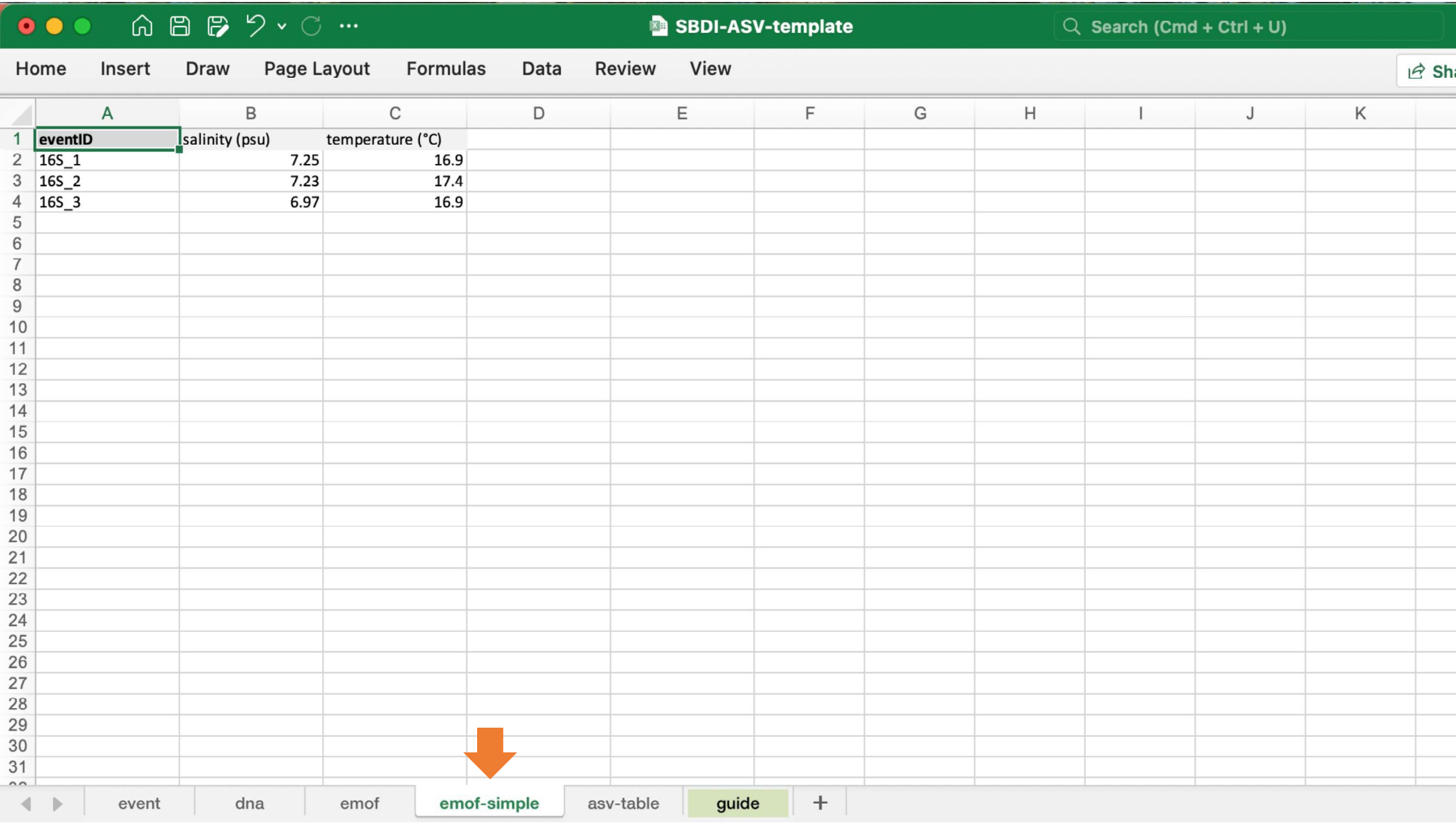

Submit data | ASV portal
asv-portal.biodiversitydata.se/submit
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOGIN
Submit data
We can help make your Amplicon Sequence Variant (ASV) dataset available to the research community via the ASV database and the Bioatlas. Read more about our platform, and the data we currently support, in the About page.
Main steps of the submission process:
1. Submit your raw data to ENA (see our guide).
2. Denoise your data to ASVs (you can use the ampliseq pipeline).
3. Download, fill in and upload the data input template (you will be instructed to login and request upload permission).
4. We will contact you to write a data-sharing contract.
5. Your data will be published in the ASV database and bioatlas.
Contact SBDI support, if you have any questions about data submission.
WHO WE ARE
Contact us
SBDI Executive Office
WHAT WE DO
History
Strategic plan
WHERE WE ARE
Sweden
The Living Atlases Community
INFORMATION
Our API:s
Accessability



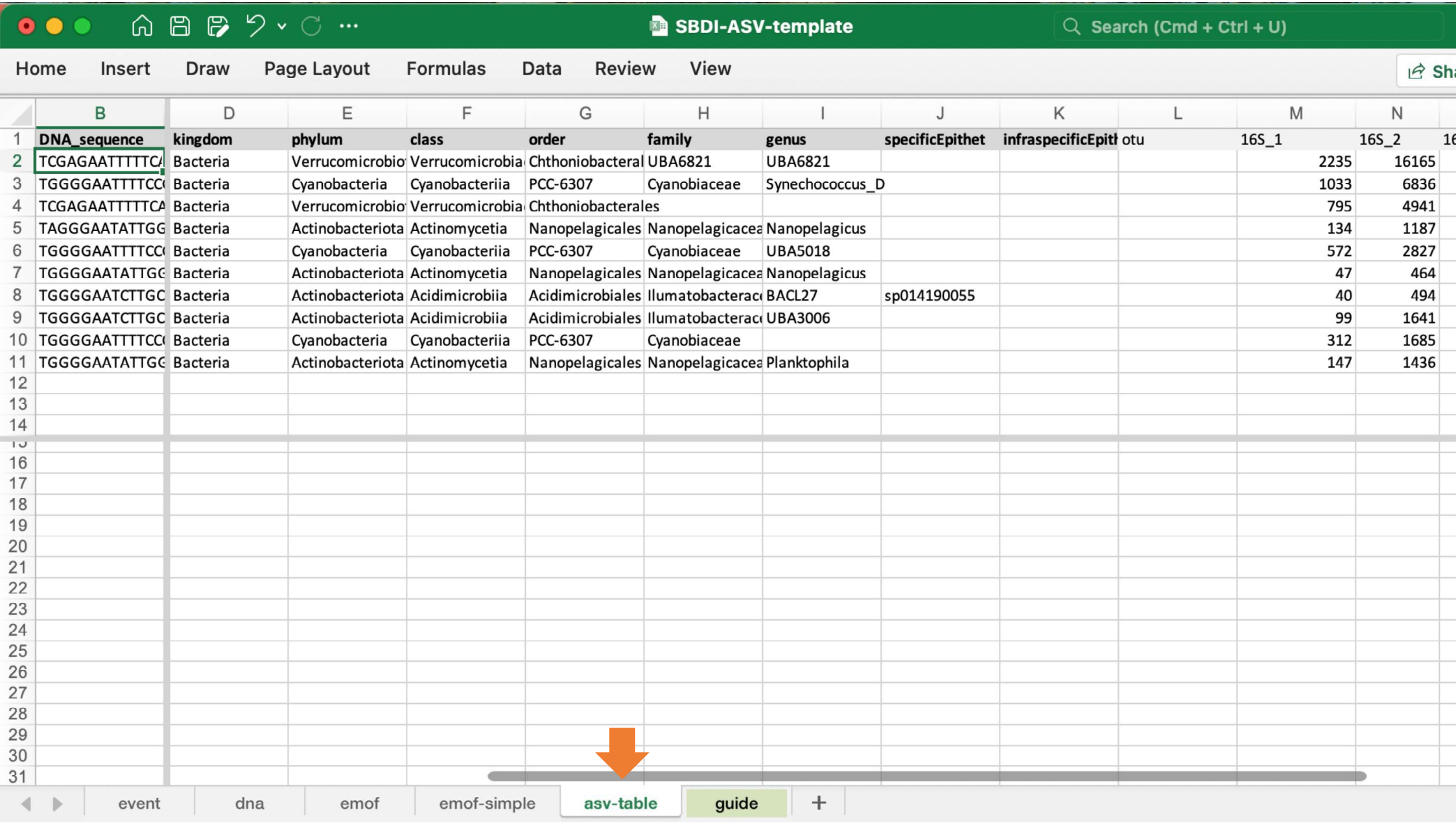
	A	B	C	D	E	F	G
1	eventID	datasetID	datasetName	institutionCode	institutionID	collectionCode	materialSampleID
2	16S_1	KTH-2013-Baltic-16S	16S data from: Diversity of Pico-	KTH	https://ror.org/026vcq606		https://www.ebi.ac.uk/ena/browser/view/SAMEA3724531
3	16S_2	KTH-2013-Baltic-16S	16S data from: Diversity of Pico-	KTH	https://ror.org/026vcq606		https://www.ebi.ac.uk/ena/browser/view/SAMEA3724535
4	16S_3	KTH-2013-Baltic-16S	16S data from: Diversity of Pico-	KTH	https://ror.org/026vcq606		https://www.ebi.ac.uk/ena/browser/view/SAMEA3724536



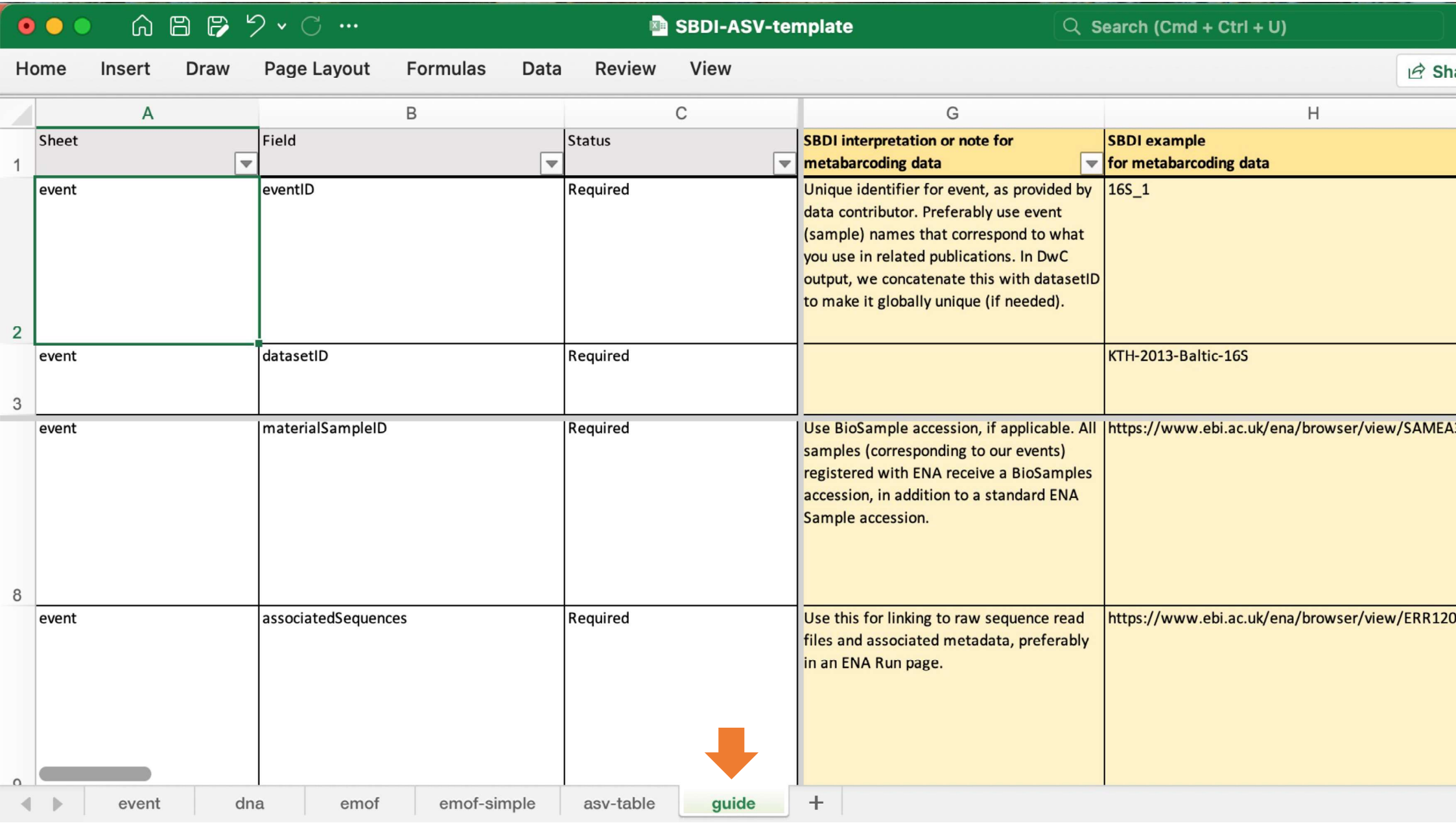
	G	H	I	J	K	L
1	pcr_primer_name_forward	pcr_primer_name_reverse	pcr_primer_forward	pcr_primer_reverse	denoising_appr	env_broad_scale
2	341F	805R	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC	DADA2	aquatic biome [ENVO:00002
3	341F	805R	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC	DADA2	aquatic biome [ENVO:00002
4	341F	805R	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATCC	DADA2	aquatic biome [ENVO:00002

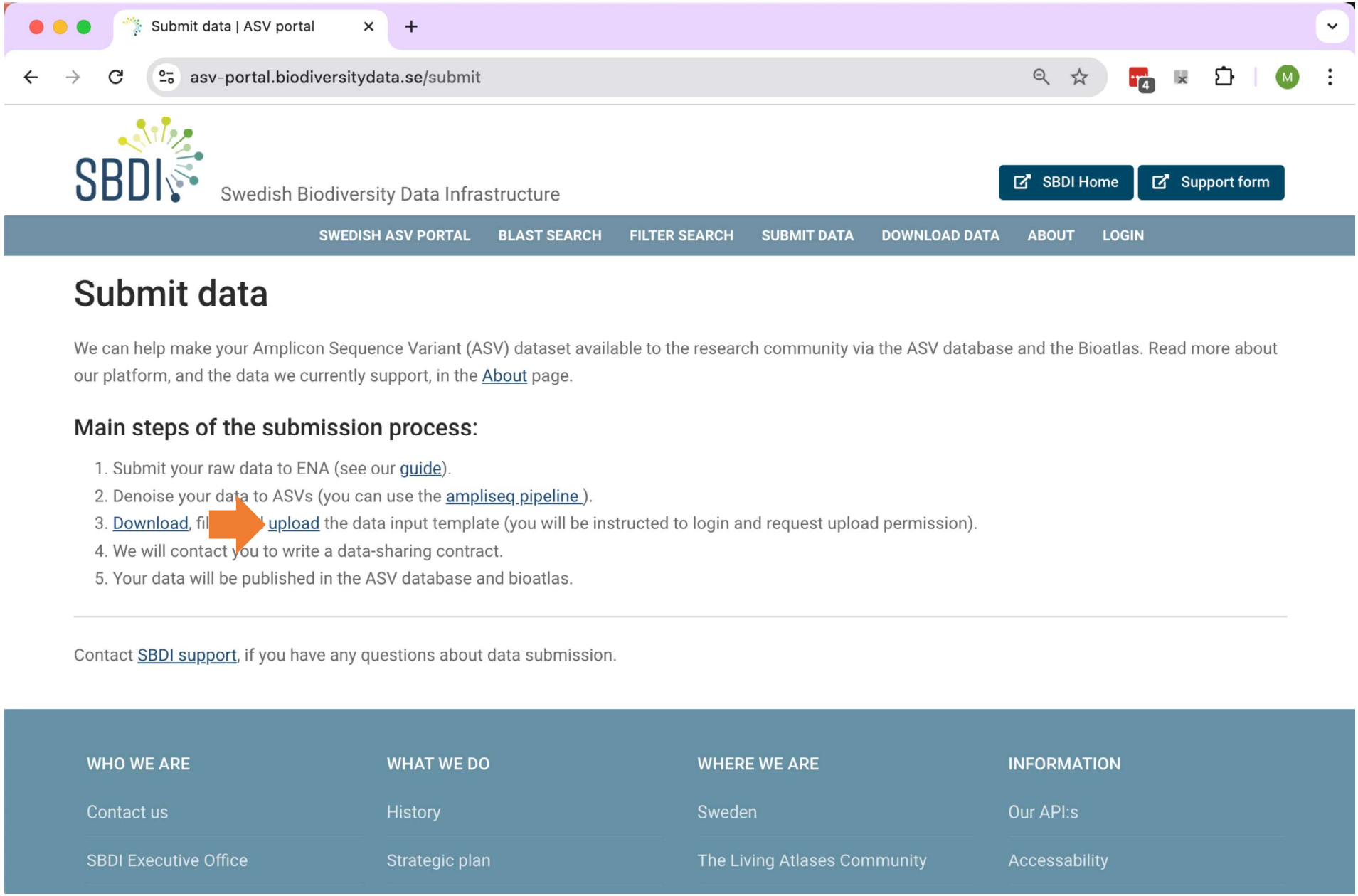


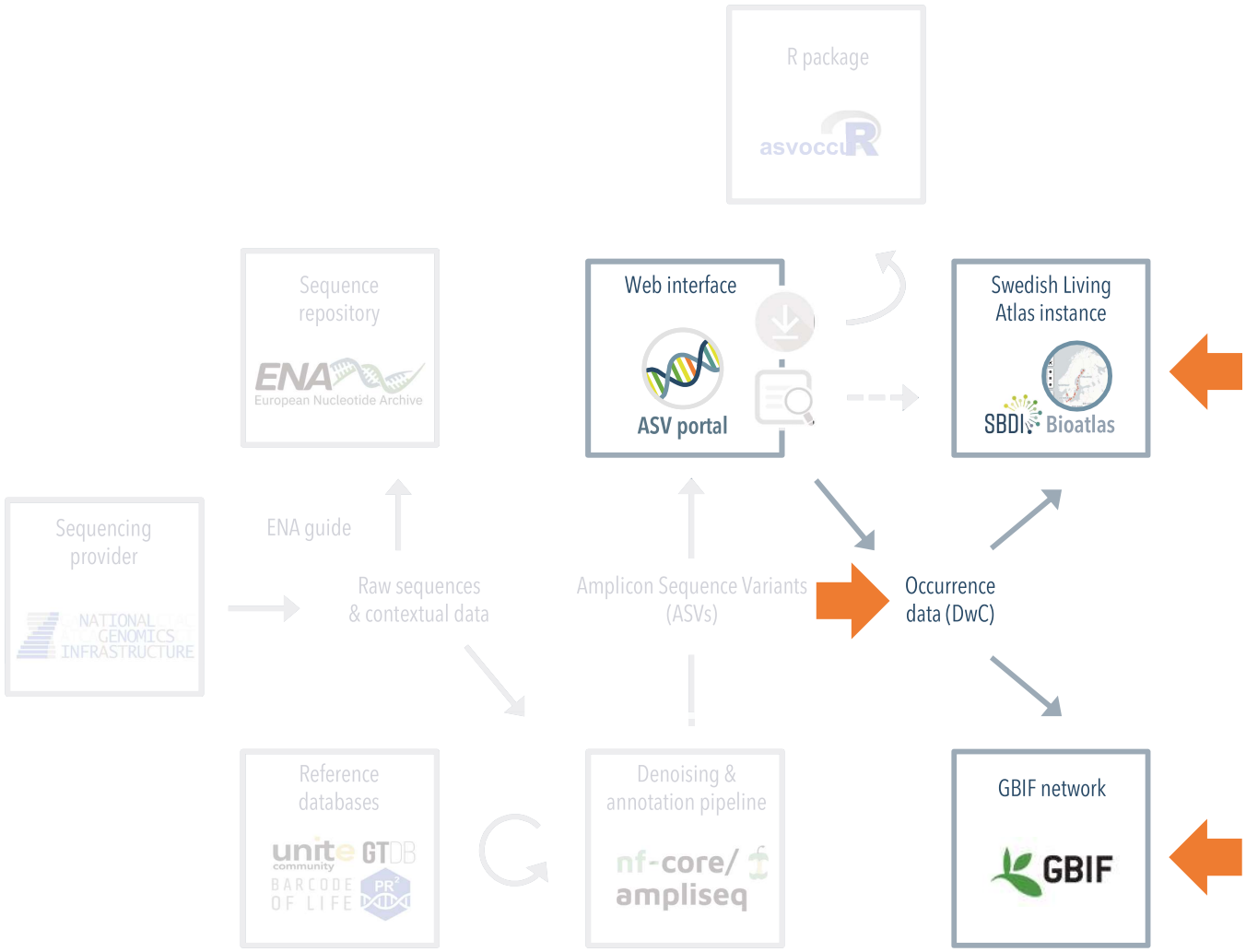
eventID	salinity (psu)	temperature (°C)
16S_1	7.25	16.9
16S_2	7.23	17.4
16S_3	6.97	16.9



	B	D	E	F	G	H	I	J	K	L	M	N
1	DNA_sequence	kingdom	phylum	class	order	family	genus	specificEpithet	infraspecificEpitl	otu	16S_1	16S_2
2	TCGAGAATTTTTCA	Bacteria	Verrucomicrobio	Verrucomicrobia	Chthoniobacteral	UBA6821	UBA6821				2235	16165
3	TGGGGAATTTTCC	Bacteria	Cyanobacteria	Cyanobacteriia	PCC-6307	Cyanobiaceae	Synechococcus_D				1033	6836
4	TCGAGAATTTTTCA	Bacteria	Verrucomicrobio	Verrucomicrobia	Chthoniobacterales						795	4941
5	TAGGGAATATTGG	Bacteria	Actinobacteriota	Actinomycetia	Nanopelagicales	Nanopelagicacea	Nanopelagicus				134	1187
6	TGGGGAATTTTCC	Bacteria	Cyanobacteria	Cyanobacteriia	PCC-6307	Cyanobiaceae	UBA5018				572	2827
7	TGGGGAATATTGG	Bacteria	Actinobacteriota	Actinomycetia	Nanopelagicales	Nanopelagicacea	Nanopelagicus				47	464
8	TGGGGAATCTTGC	Bacteria	Actinobacteriota	Acidimicrobiia	Acidimicrobiales	Ilumatobacterac	BACL27	sp014190055			40	494
9	TGGGGAATCTTGC	Bacteria	Actinobacteriota	Acidimicrobiia	Acidimicrobiales	Ilumatobacterac	UBA3006				99	1641
10	TGGGGAATTTTCC	Bacteria	Cyanobacteria	Cyanobacteriia	PCC-6307	Cyanobiaceae					312	1685
11	TGGGGAATATTGG	Bacteria	Actinobacteriota	Actinomycetia	Nanopelagicales	Nanopelagicacea	Planktophila				147	1436
12												
13												
14												
16												
17												
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
31												


SBDI-ASV-template
Search (Cmd + Ctrl + U)
Home
Insert
Draw
Page Layout
Formulas
Data
Review
View
Sha
A
B
C
G
H
Sheet
Field
Status
SBDI interpretation or note for metabarcoding data
SBDI example for metabarcoding data
event
eventID
Required
Unique identifier for event, as provided by data contributor. Preferably use event (sample) names that correspond to what you use in related publications. In DwC output, we concatenate this with datasetID to make it globally unique (if needed).
16S_1
event
datasetID
Required
KTH-2013-Baltic-16S
event
materialSampleID
Required
Use BioSample accession, if applicable. All samples (corresponding to our events) registered with ENA receive a BioSamples accession, in addition to a standard ENA Sample accession.
https://www.ebi.ac.uk/ena/browser/view/SAMEA
event
associatedSequences
Required
Use this for linking to raw sequence read files and associated metadata, preferably in an ENA Run page.
https://www.ebi.ac.uk/ena/browser/view/ERR120
event
dna
emof
emof-simple
asv-table
guide


Submit data | ASV portal
asv-portal.biodiversitydata.se/submit
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOGIN
Submit data
We can help make your Amplicon Sequence Variant (ASV) dataset available to the research community via the ASV database and the Bioatlas. Read more about our platform, and the data we currently support, in the About page.
Main steps of the submission process:
1. Submit your raw data to ENA (see our guide).
2. Denoise your data to ASVs (you can use the ampliseq pipeline).
3. Download, fi upload the data input template (you will be instructed to login and request upload permission).
4. We will contact you to write a data-sharing contract.
5. Your data will be published in the ASV database and bioatlas.
Contact SBDI support, if you have any questions about data submission.
WHO WE ARE
Contact us
SBDI Executive Office
WHAT WE DO
History
Strategic plan
WHERE WE ARE
Sweden
The Living Atlases Community
INFORMATION
Our API:s
Accessability


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)
Reference databases
unite community
GTDB
BARCODE OF LIFE
PR2
Denoising & annotation pipeline
nf-core/ampliseq
GBIF network
GBIF

SBDI Swedish Biodiversity Data Infrastructure

Search species, datas

Login Sign up

SEARCH AND EXPLORE SBDI TOOLS HELP

Narrow your results

Occurrence

Year

☐ 2013 (11,440)

Miscellaneous

Record type

☐ Material Sample (11,440)

Scientific name

☐ Cyanobacteriales (290)

☐ Cyanobacteriia (284)

☐ Flavobacterium (279)

☐ Bacteria (230)

choose more...

Institution

☐ KTH Royal Institute of Technology (11,440)

Collection

☐ 16S data from: Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea (Hu et al. 2016) (11,440)

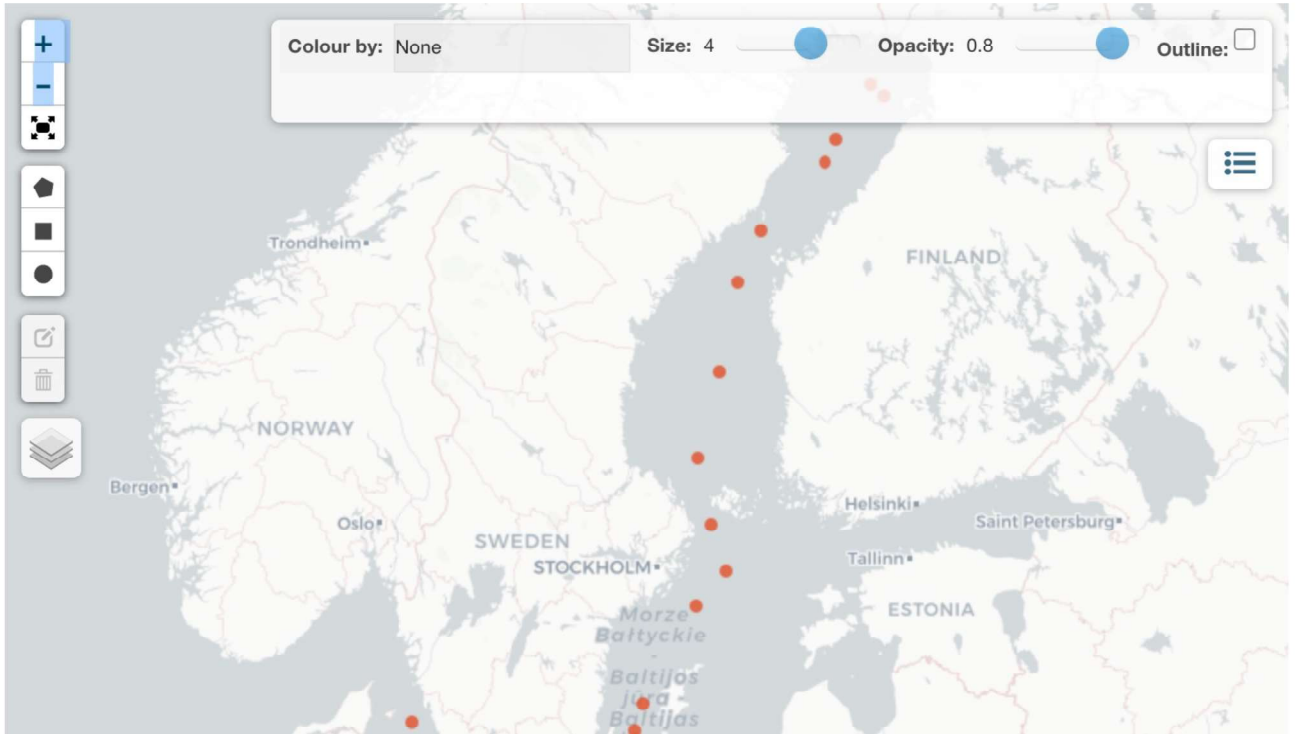
Data resource

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

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

Records | Map | Charts

View in spatial portal | Download map



Get data How-to Tools Community About maria.prager

GBIF

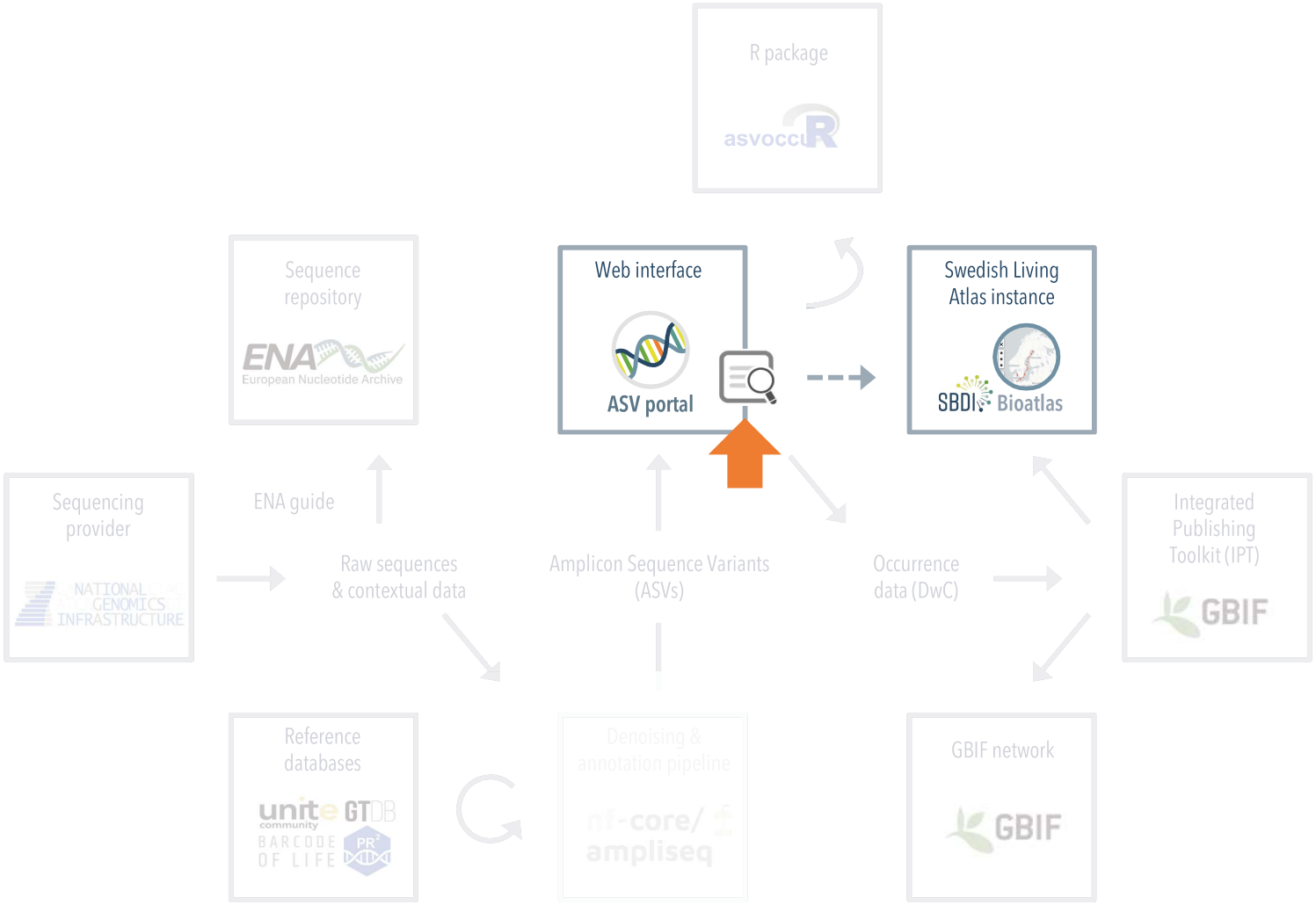
OCCURRENCE DATASET | REGISTERED SEPTEMBER 22, 2021

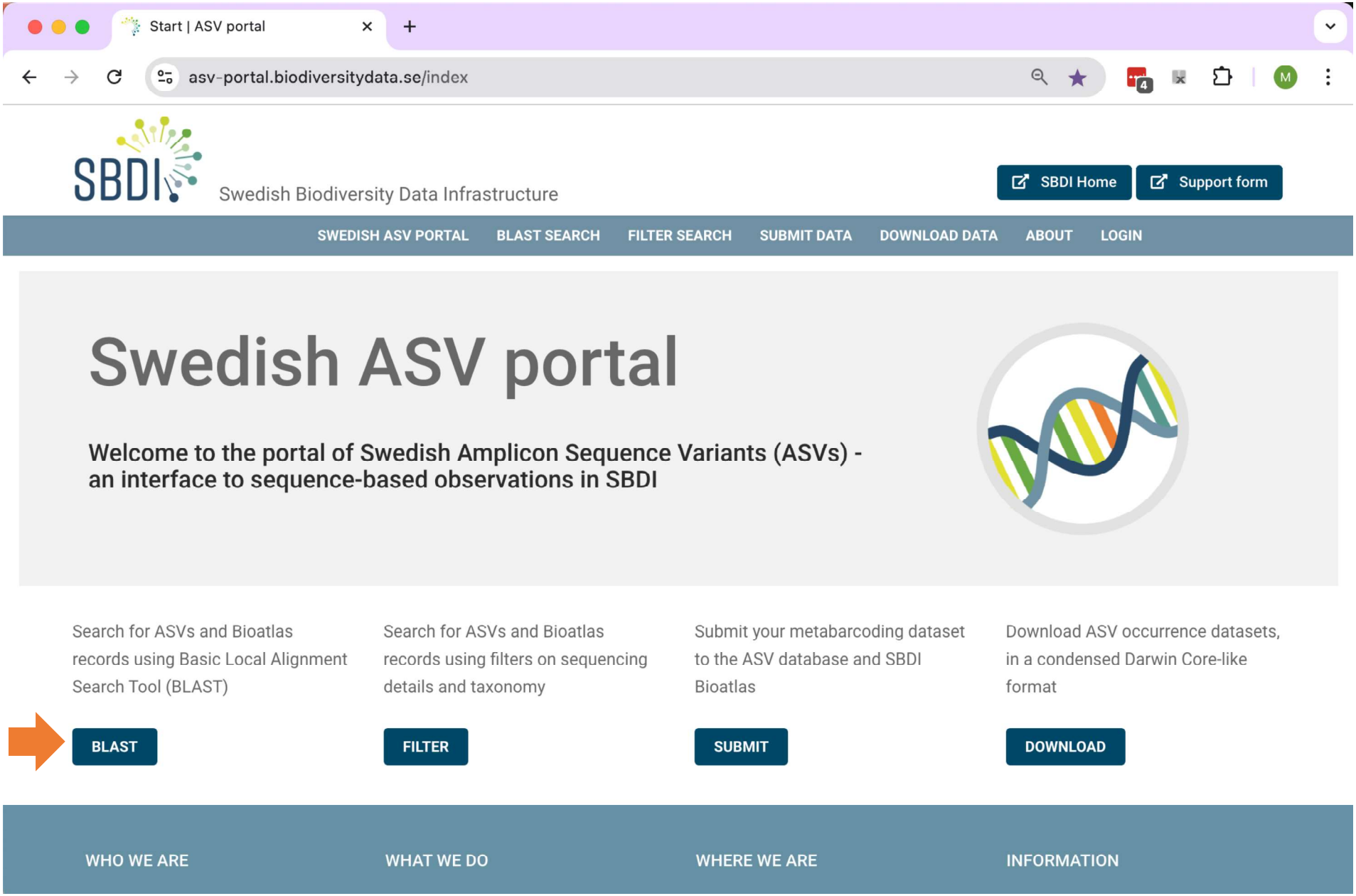
16S data from: Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea (Hu et al. 2016)

Published by KTH Royal Institute of Technology

11,440	100%	100%	100%
Occurrences	With taxon match	With coordinates	With year

11,440 GEOREFERENCED RECORDS


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
GBIF network
GBIF


Start | ASV portal
asv-portal.biodiversitydata.se/index
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOGIN
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)
BLAST
Search for ASVs and Bioatlas records using filters on sequencing details and taxonomy
FILTER
Submit your metabarcoding dataset to the ASV database and SBDI Bioatlas
SUBMIT
Download ASV occurrence datasets, in a condensed Darwin Core-like format
DOWNLOAD
WHO WE ARE
WHAT WE DO
WHERE WE ARE
INFORMATION

Query sequence(s)

Nucleotide sequence(s), in fasta format, to compare against ASVs (subject sequences) in reference database 830/50000 characters

```
TGGGGAATTTTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGGAGGATGAAGTCCCTTGGGACGTAAACTCCTTTCGACCGGGACGATTATGACGGTACCGGTGGA
AGAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGGTAATACGAGGGGGGCAAGCGTTGTTCGGAATTATTGGGCGTAAAGGGCGCGTAGGCGGTGCGGTAAGTCACCTGTGAA
ACCTCTGGGCTCAACCCAGAGCCTGCAGGCGAAACTGCCGTGCTGGAGTATGGGAGAGGTGCGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCT
GTGGCGAAAGCGGCGCACTGGACCATAACTGACGCTGAGGCGCGAAAGCTAGGGGAGCAAACA
>test-seq-2
TGGGGAATTTTGCGCAATGGGGGAAACCCTGACGCAGCAACGCCGCGTGGAGGATGAAGCCCCTTGGGGTGTAAACTCCTTTCGATCGGGACGATTATGACGGTACCGGATGA
AGAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGGGTGCAAGCGTTGTTCGGAATTATTGGGCGTAAAGGGTGCGTAGGCGGTGCGGTAAGTCTTTTGTGAAA
TCTCCGGGCTCAACCCGGAGCCTGCAAGGGAAACTGCCGTGCTTGAGTGTGGGAGAGGTGAGTGGAATTCCCGGTGTAGCGGTGAAATGCGTAGATATCGGGAGGAACACCTG
TGGCGAAAGCGGCTCACTGGACCACAACTGACGCTGATGCACGAAAGCTAGGGGAGCAAACA
```

Minimum identity (Id %)

Share of exact matches in alignment

98

Minimum query coverage (Cov %)

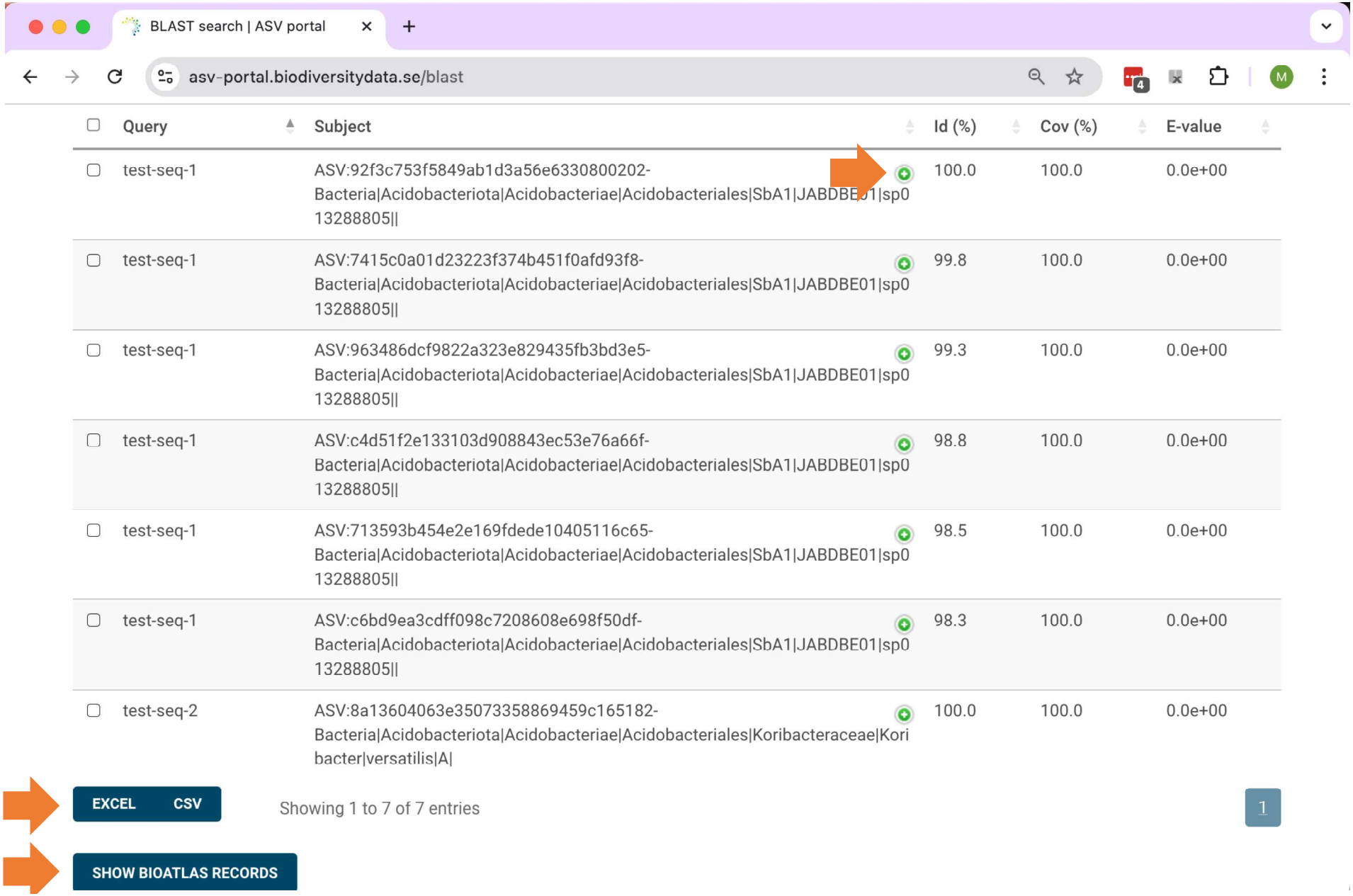
Share of aligned query bases

100

What does this mean? This corresponds to a nucleotide BLAST (*BLASTn*) search filtered on *-perc_identity* and *-qcov_hsp_perc*. Query cover is calculated per *High Scoring Pair (HSP)*, i.e. per ungapped local pairwise alignment between query and subject, and we only report the best HSP for each query-subject hit. Result fields correspond to *qacc*, *stitle*, *pident*, *qcovhsp* and *evalue* in tabular BLASTn output format (*–outfmt=6*), and rows are sorted on *Query* and *E-value* by default. E-value specifies the number of hits of similar (or higher) score that we expect to see by chance, i.e. smaller values indicate 'better' hits.

BLAST CLEAR

Tip: You may have to **scroll down and wait** for results to load. Then tick a checkbox to (de)select a row, toggle +/- symbols to show/hide sequences, and click bottom *Show Bioatlas records* button to show occurrences of selected ASV:s in the Bioatlas. You can also download search results as Excel/CSV, and use *taxonID* column to link these to Bioatlas records.


BLAST search | ASV portal
asv-portal.biodiversitydata.se/blast
Query
Subject
Id (%)
Cov (%)
E-value
test-seq-1
ASV:92f3c753f5849ab1d3a56e6330800202-Bacteria|Acidobacteriota|Acidobacteriae|Acidobacteriales|SbA1|JABDBE01|sp013288805||
100.0
100.0
0.0e+00
test-seq-1
ASV:7415c0a01d23223f374b451f0afd93f8-Bacteria|Acidobacteriota|Acidobacteriae|Acidobacteriales|SbA1|JABDBE01|sp013288805||
99.8
100.0
0.0e+00
test-seq-1
ASV:963486dcf9822a323e829435fb3bd3e5-Bacteria|Acidobacteriota|Acidobacteriae|Acidobacteriales|SbA1|JABDBE01|sp013288805||
99.3
100.0
0.0e+00
test-seq-1
ASV:c4d51f2e133103d908843ec53e76a66f-Bacteria|Acidobacteriota|Acidobacteriae|Acidobacteriales|SbA1|JABDBE01|sp013288805||
98.8
100.0
0.0e+00
test-seq-1
ASV:713593b454e2e169fdede10405116c65-Bacteria|Acidobacteriota|Acidobacteriae|Acidobacteriales|SbA1|JABDBE01|sp013288805||
98.5
100.0
0.0e+00
test-seq-1
ASV:c6bd9ea3cdff098c7208608e698f50df-Bacteria|Acidobacteriota|Acidobacteriae|Acidobacteriales|SbA1|JABDBE01|sp013288805||
98.3
100.0
0.0e+00
test-seq-2
ASV:8a13604063e35073358869459c165182-Bacteria|Acidobacteriota|Acidobacteriae|Acidobacteriales|Koribacteraceae|Koribacter|versatilis|A|
100.0
100.0
0.0e+00
EXCEL
CSV
Showing 1 to 7 of 7 entries
1
SHOW BIOATLAS RECORDS

Start | ASV portal

asv-portal.biodiversitydata.se/index

SBDI Swedish Biodiversity Data Infrastructure

SBDI Home | Support form

SWEDISH ASV PORTAL | BLAST SEARCH | FILTER SEARCH | SUBMIT DATA | DOWNLOAD DATA | ABOUT | LOGIN

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)

BLAST

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

FILTER

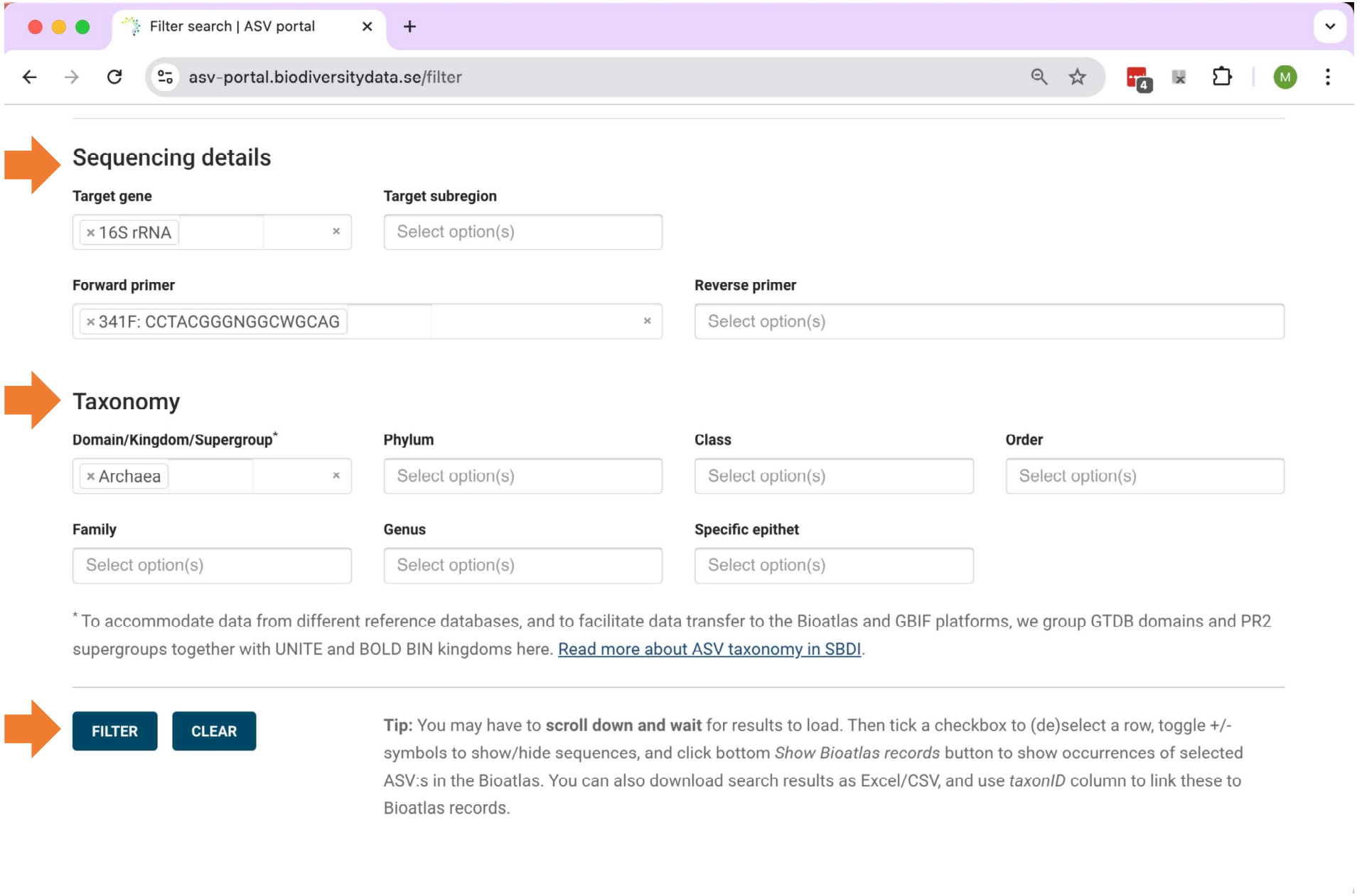
Submit your metabarcoding dataset to the ASV database and SBDI Bioatlas

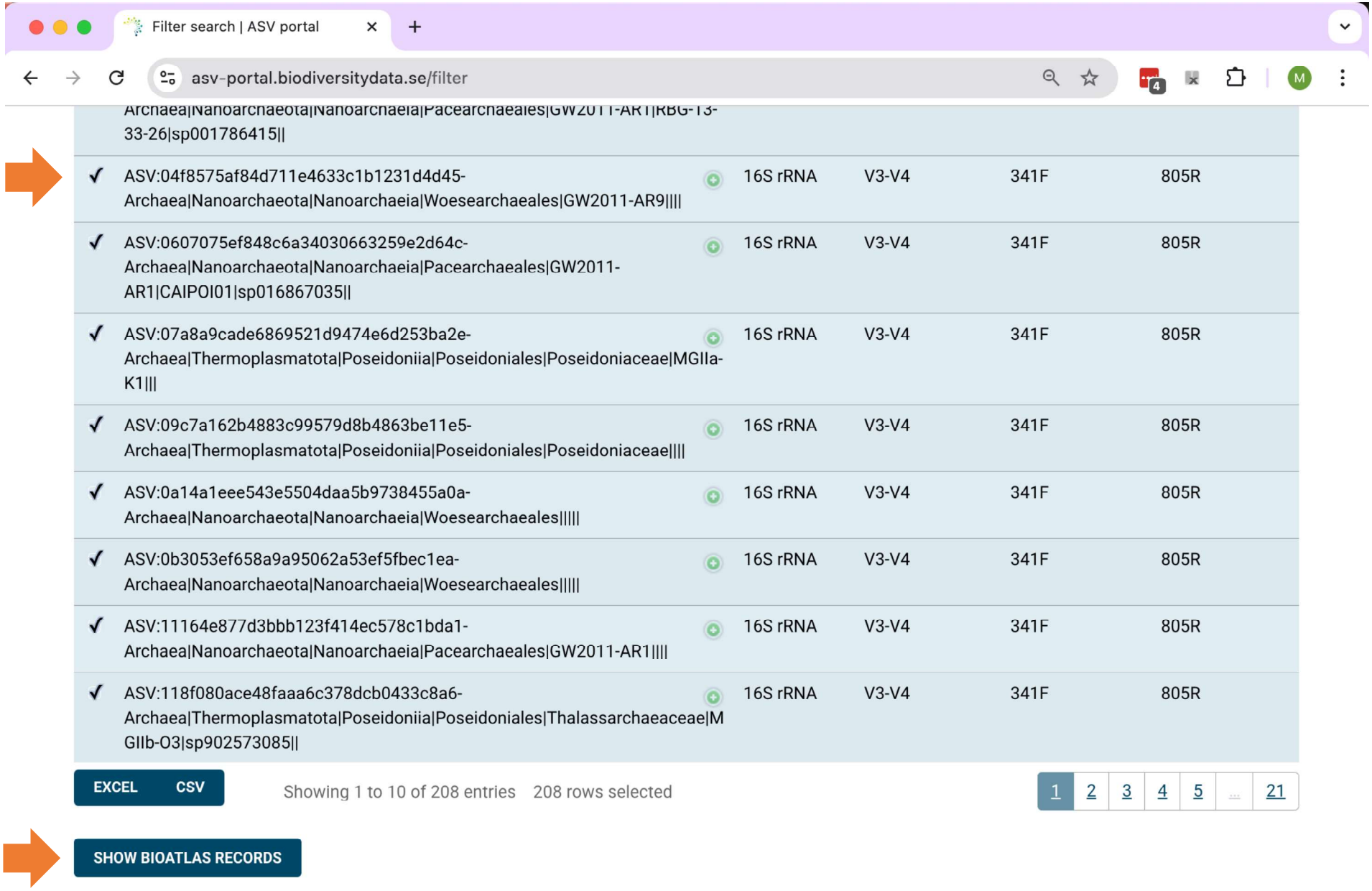
SUBMIT

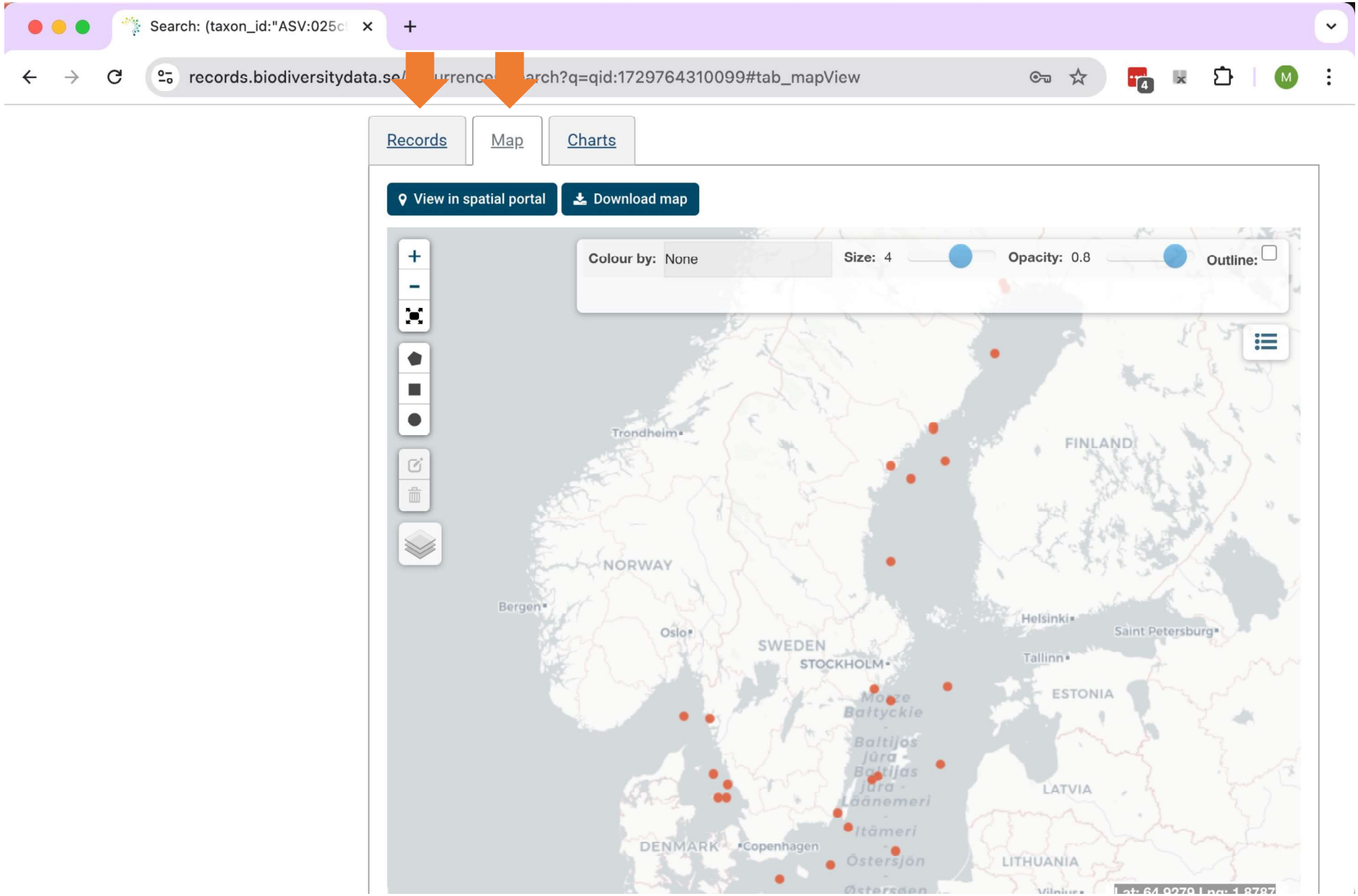
Download ASV occurrence datasets, in a condensed Darwin Core-like format

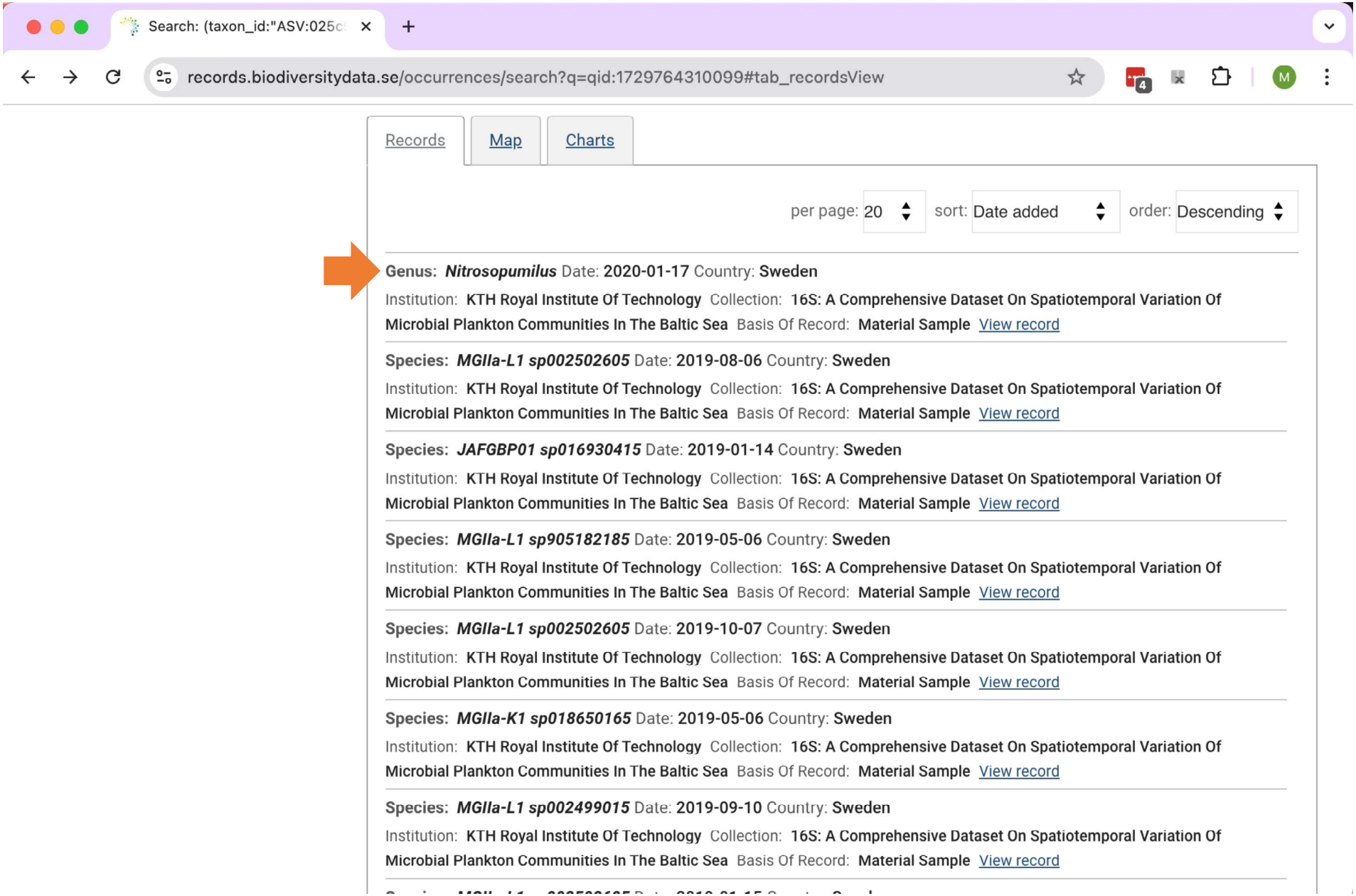
DOWNLOAD

WHO WE ARE | WHAT WE DO | WHERE WE ARE | INFORMATION


Filter search | ASV portal
asv-portal.biodiversitydata.se/filter
Sequencing details
Target gene
× 16S rRNA
Target subregion
Select option(s)
Forward primer
× 341F: CCTACGGGNGGCWGCAG
Reverse primer
Select option(s)
Taxonomy
Domain/Kingdom/Supergroup*
× Archaea
Phylum
Select option(s)
Class
Select option(s)
Order
Select option(s)
Family
Select option(s)
Genus
Select option(s)
Specific epithet
Select option(s)
* To accommodate data from different reference databases, and to facilitate data transfer to the Bioatlas and GBIF platforms, we group GTDB domains and PR2 supergroups together with UNITE and BOLD BIN kingdoms here. Read more about ASV taxonomy in SBDI.
FILTER
CLEAR
Tip: You may have to scroll down and wait for results to load. Then tick a checkbox to (de)select a row, toggle +/- symbols to show/hide sequences, and click bottom Show Bioatlas records button to show occurrences of selected ASV:s in the Bioatlas. You can also download search results as Excel/CSV, and use taxonID column to link these to Bioatlas records.

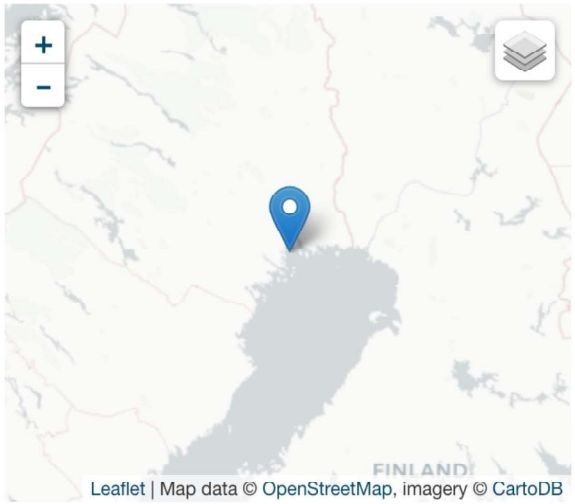

Filter search | ASV portal
asv-portal.biodiversitydata.se/filter
Archaea|Nanoarchaeota|Nanoarchaeia|Pacearchaeales|GW2011-AR1|RBG-13-33-26|sp001786415||
ASV:04f8575af84d711e4633c1b1231d4d45-Archaea|Nanoarchaeota|Nanoarchaeia|Woesearchaeales|GW2011-AR9|||| 16S rRNA V3-V4 341F 805R
ASV:0607075ef848c6a34030663259e2d64c-Archaea|Nanoarchaeota|Nanoarchaeia|Pacearchaeales|GW2011-AR1|CAIPOI01|sp016867035|| 16S rRNA V3-V4 341F 805R
ASV:07a8a9cade6869521d9474e6d253ba2e-Archaea|Thermoplasmatota|Poseidoniia|Poseidoniales|Poseidoniaceae|MGIIa-K1||| 16S rRNA V3-V4 341F 805R
ASV:09c7a162b4883c99579d8b4863be11e5-Archaea|Thermoplasmatota|Poseidoniia|Poseidoniales|Poseidoniaceae|||| 16S rRNA V3-V4 341F 805R
ASV:0a14a1eee543e5504daa5b9738455a0a-Archaea|Nanoarchaeota|Nanoarchaeia|Woesearchaeales||||| 16S rRNA V3-V4 341F 805R
ASV:0b3053ef658a9a95062a53ef5fbec1ea-Archaea|Nanoarchaeota|Nanoarchaeia|Woesearchaeales||||| 16S rRNA V3-V4 341F 805R
ASV:11164e877d3bbb123f414ec578c1bda1-Archaea|Nanoarchaeota|Nanoarchaeia|Pacearchaeales|GW2011-AR1|||| 16S rRNA V3-V4 341F 805R
ASV:118f080ace48faaa6c378dcb0433c8a6-Archaea|Thermoplasmatota|Poseidoniia|Poseidoniales|Thalassarchaeaceae|MGIIb-O3|sp902573085|| 16S rRNA V3-V4 341F 805R
EXCEL
CSV
Showing 1 to 10 of 208 entries
208 rows selected
1
2
3
4
5
...
21
SHOW BIOATLAS RECORDS


Search: (taxon_id:"ASV:025c
records.biodiversitydata.se/
rch?q=qid:1729764310099#tab_mapView
Records
Map
Charts
View in spatial portal
Download map
Colour by: None
Size: 4
Opacity: 0.8
Outline:
Trondheim
FINLAND
NORWAY
Bergen
Oslo
SWEDEN
STOCKHOLM
Helsinki
Saint Petersburg
Tallinn
ESTONIA
LATVIA
LITHUANIA
DENMARK
Copenhagen


Search: (taxon_id:"ASV:025c
records.biodiversitydata.se/occurrences/search?q=qid:1729764310099#tab_recordsView
Records
Map
Charts
per page: 20
sort: Date added
order: Descending
Genus: Nitrosopumilus Date: 2020-01-17 Country: Sweden
Institution: KTH Royal Institute Of Technology Collection: 16S: A Comprehensive Dataset On Spatiotemporal Variation Of Microbial Plankton Communities In The Baltic Sea Basis Of Record: Material Sample View record
Species: MGIIa-L1 sp002502605 Date: 2019-08-06 Country: Sweden
Institution: KTH Royal Institute Of Technology Collection: 16S: A Comprehensive Dataset On Spatiotemporal Variation Of Microbial Plankton Communities In The Baltic Sea Basis Of Record: Material Sample View record
Species: JAFGBP01 sp016930415 Date: 2019-01-14 Country: Sweden
Institution: KTH Royal Institute Of Technology Collection: 16S: A Comprehensive Dataset On Spatiotemporal Variation Of Microbial Plankton Communities In The Baltic Sea Basis Of Record: Material Sample View record
Species: MGIIa-L1 sp905182185 Date: 2019-05-06 Country: Sweden
Institution: KTH Royal Institute Of Technology Collection: 16S: A Comprehensive Dataset On Spatiotemporal Variation Of Microbial Plankton Communities In The Baltic Sea Basis Of Record: Material Sample View record
Species: MGIIa-L1 sp002502605 Date: 2019-10-07 Country: Sweden
Institution: KTH Royal Institute Of Technology Collection: 16S: A Comprehensive Dataset On Spatiotemporal Variation Of Microbial Plankton Communities In The Baltic Sea Basis Of Record: Material Sample View record
Species: MGIIa-K1 sp018650165 Date: 2019-05-06 Country: Sweden
Institution: KTH Royal Institute Of Technology Collection: 16S: A Comprehensive Dataset On Spatiotemporal Variation Of Microbial Plankton Communities In The Baltic Sea Basis Of Record: Material Sample View record
Species: MGIIa-L1 sp002499015 Date: 2019-09-10 Country: Sweden
Institution: KTH Royal Institute Of Technology Collection: 16S: A Comprehensive Dataset On Spatiotemporal Variation Of Microbial Plankton Communities In The Baltic Sea Basis Of Record: Material Sample View record

Record: PRJEB55296-16S:P2

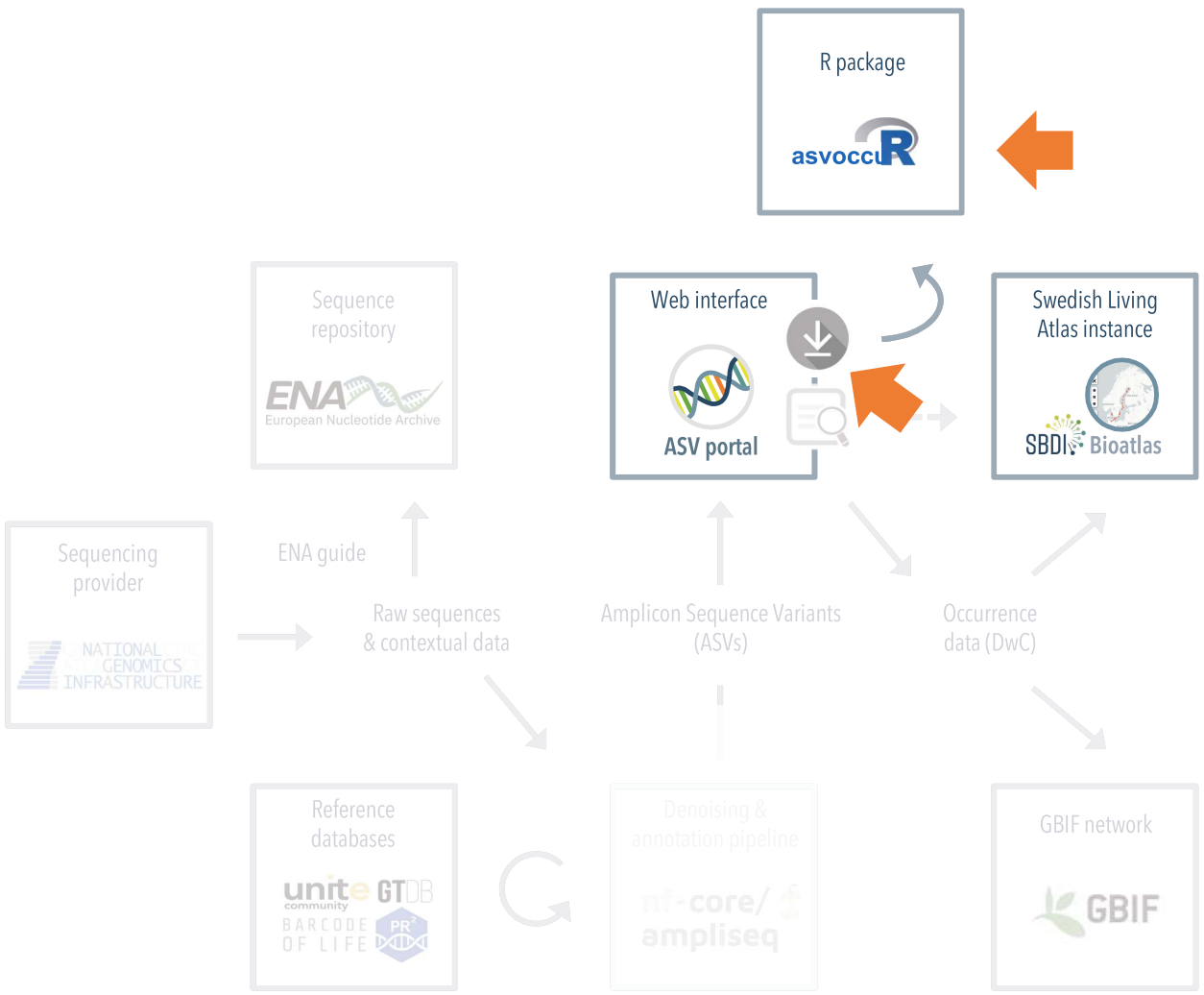
records.biodiversitydata.se/occurrences/7b6f5e31-de0d-44ca-8e65-f45c3be0e980

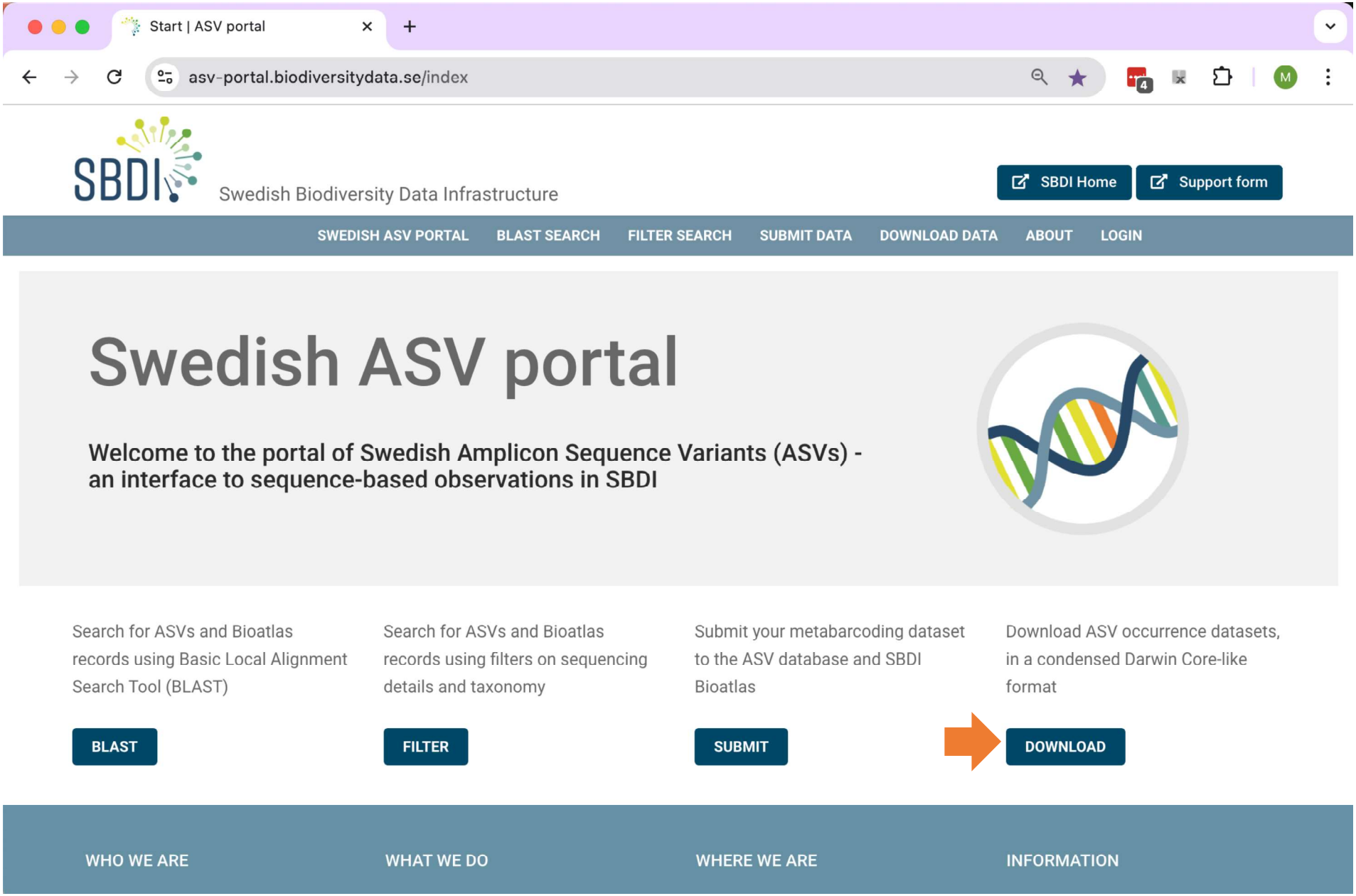


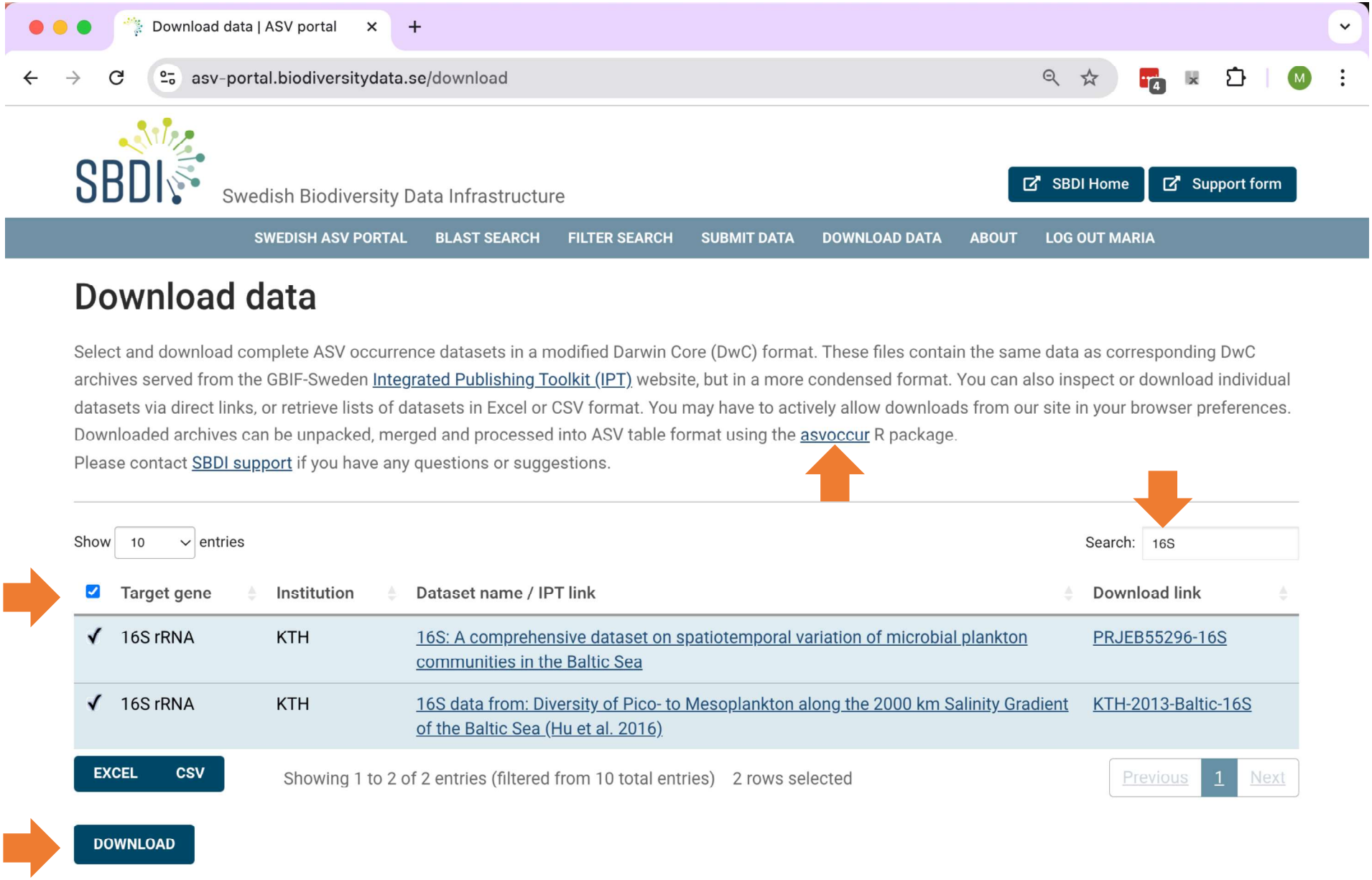
Date loaded: 2024-09-26

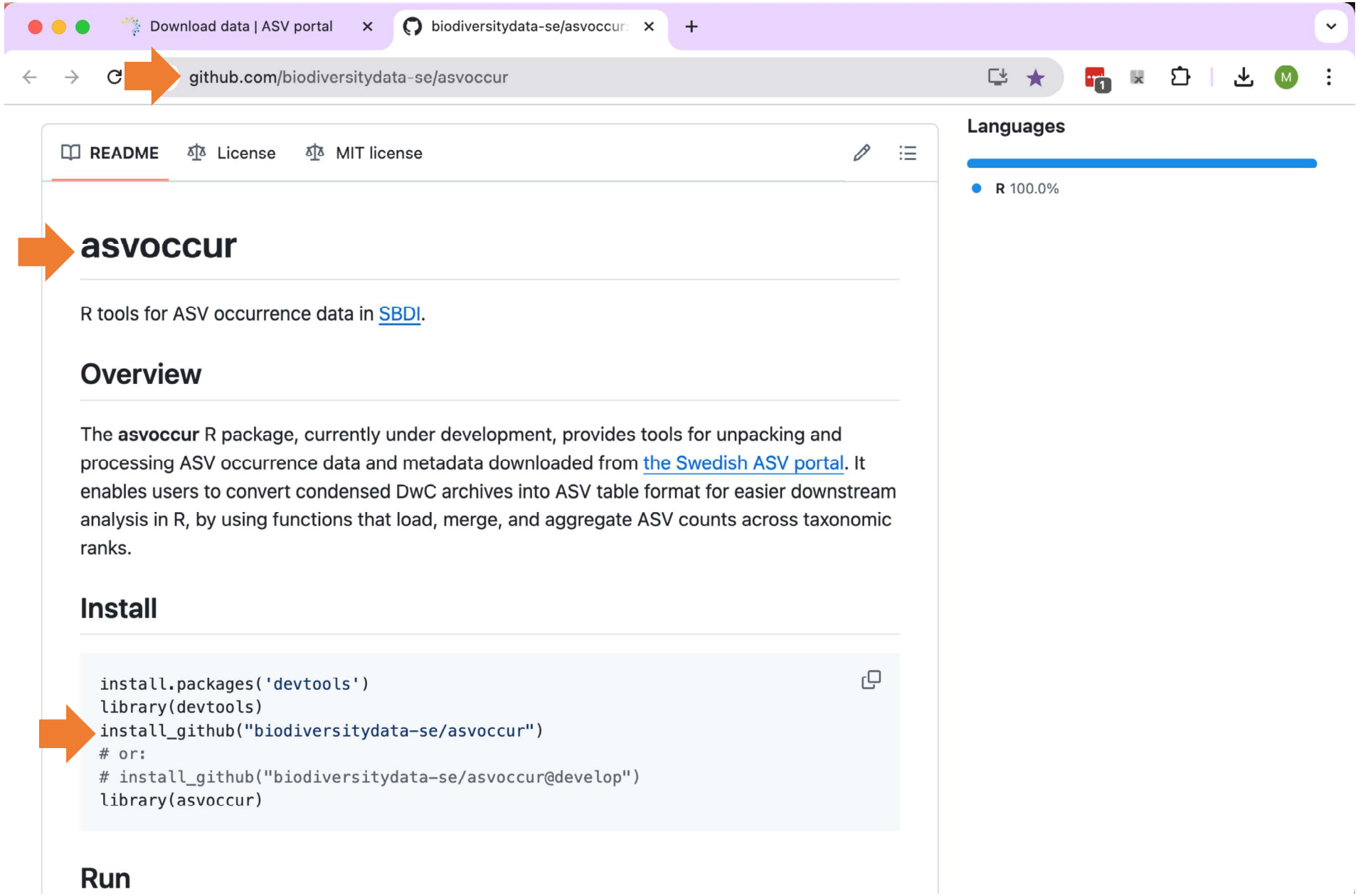
Date last processed: 2024-09-26

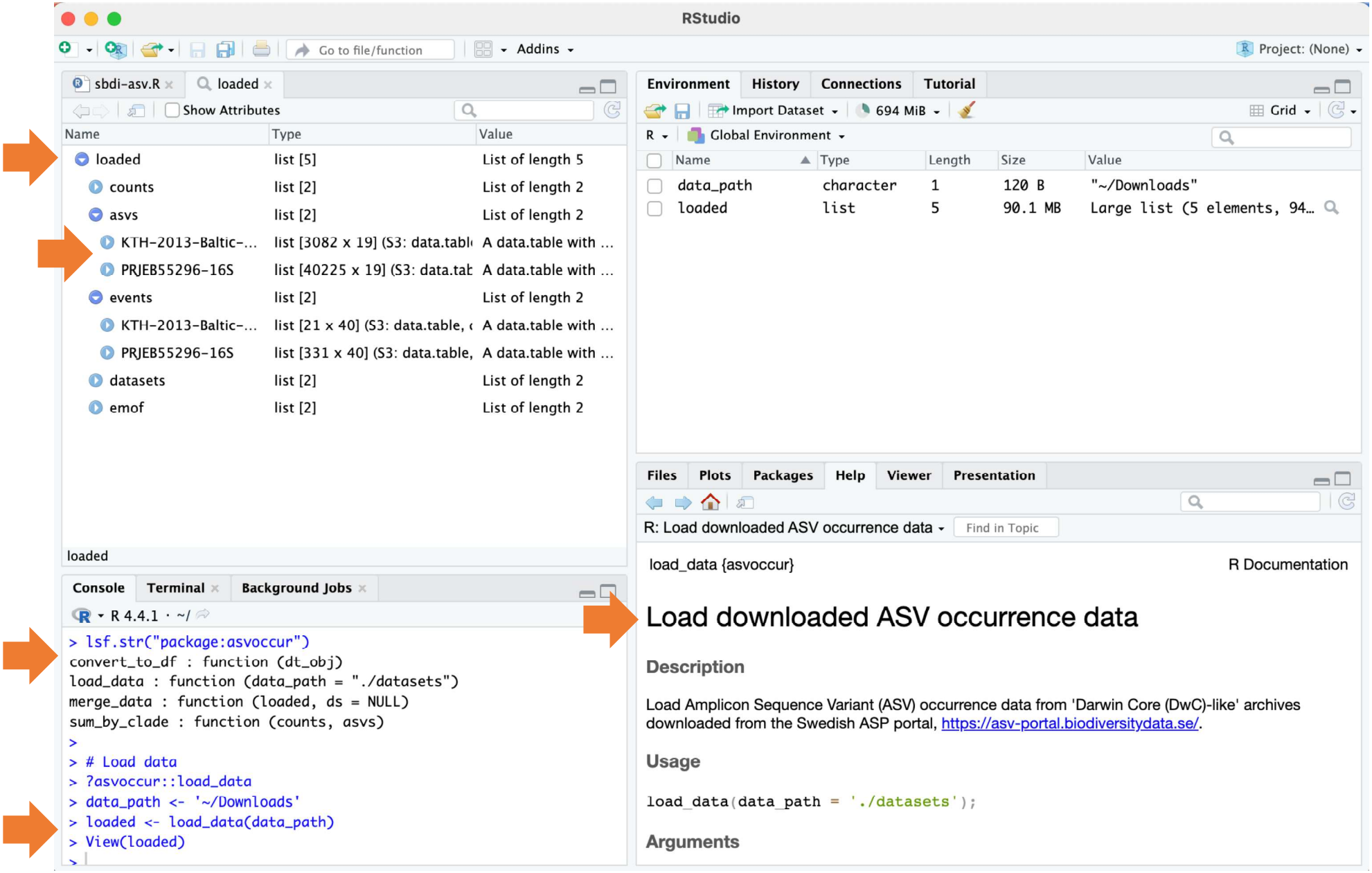
Collector	KTH Royal Institute of Technology | Swedish Meteorological and Hydrological Institute (SMHI) | Umeå University (UMU) *Supplied as ""*
License	CC0
Institution ID	https://ror.org/026vcq606
Presence/Absence	PRESENT
Organism quantity	28
Material sample ID	https://www.ebi.ac.uk/ena/browser/view/SAMEA110701943
Organism quantity type	DNA sequence reads
Sampling protocol	[https://dx.doi.org/10.17504/protocols.io.bucjnsun]
Dataset / Survey name	[16S: A comprehensive dataset on spatiotemporal variation of microbial plankton communities in the Baltic Sea]
Previous identifications	By data provider: Archaea | Thermoproteota | Nitrososphaeria | Nitrososphaerales | Nitrosopumilaceae | Nitrosopumilus | | | ; By ASV portal: Archaea | Thermoproteota | Nitrososphaerales | Nitrososphaeria | Nitrosopumilaceae | Nitrosopumilus | sp013203245 | |
Dataset	[PRJEB55296-16S]
Associated sequences	https://www.ebi.ac.uk/ena/browser/view/ERR10113559
Identification references	https://docs.biodiversitydata.se/analyse-data/molecular-tools/#taxonomy-annotation
Date identified	2023-06-16
Identification remarks	Ampliseq v2.5.0 (https://nf-co.re/ampliseq) DADA2:assignTaxonomy annotation against SBDI-GTDB-R07-RS207-1 (https://scilifelab.figshare.com/articles/dataset/SBDI_Sativa_curated_16S_GTDB_database/14869077/4); confidence at lowest specified (ASV portal) taxon: 0.99

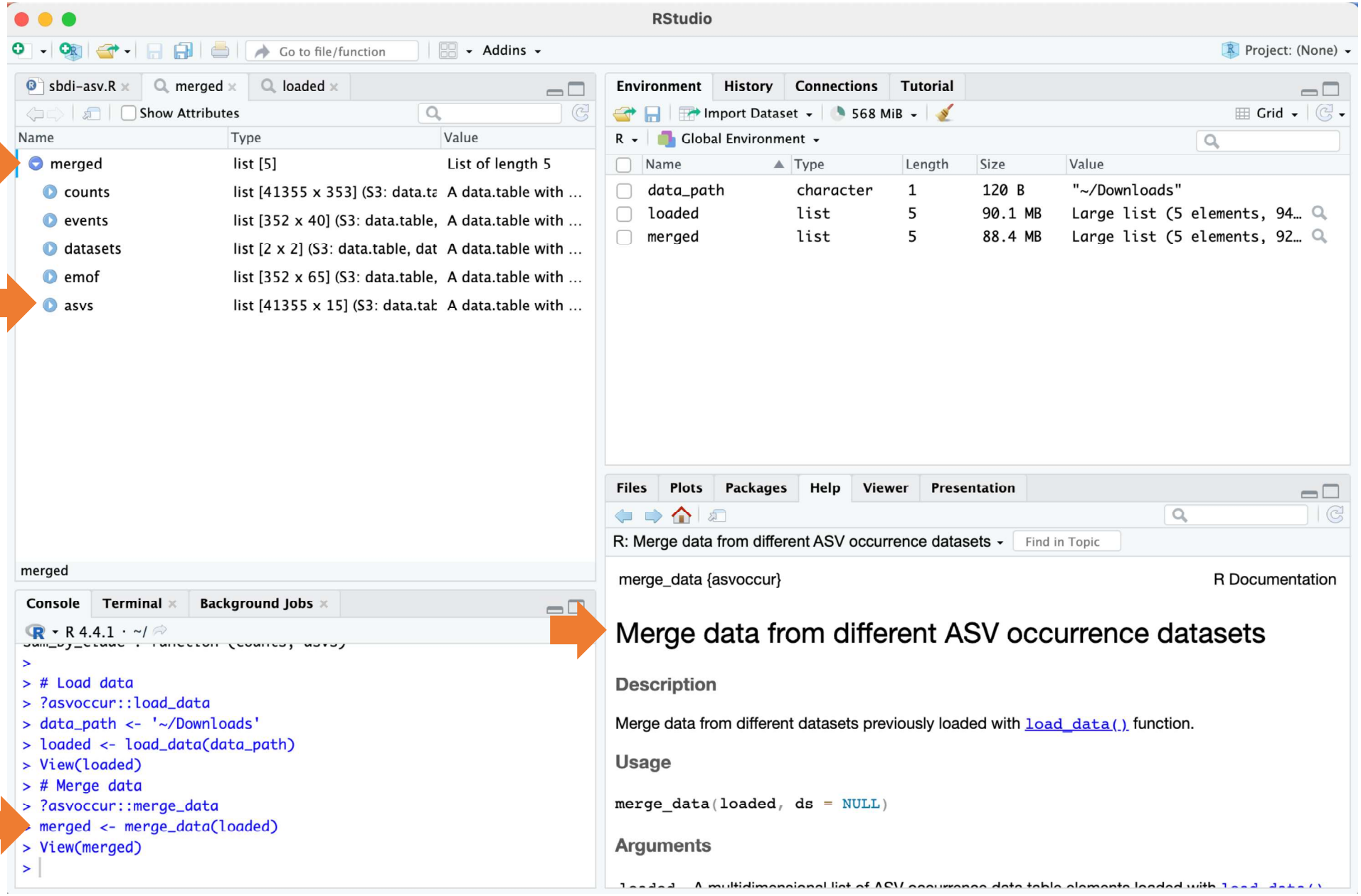

R package
asvoccuR
Sequence repository
ENA
European Nucleotide Archive
Web interface
ASV portal
Swedish Living Atlas instance
SBDI Bioatlas
Sequencing provider
NATIONAL GENOMICS INFRASTRUCTURE
ENA guide
Raw sequences & contextual data
Amplicon Sequence Variants (ASVs)
Occurrence data (DwC)
Reference databases
unite community
GTDB
BARCODE OF LIFE
Denoising & annotation pipeline
nf-core/ampliseq
GBIF network
GBIF

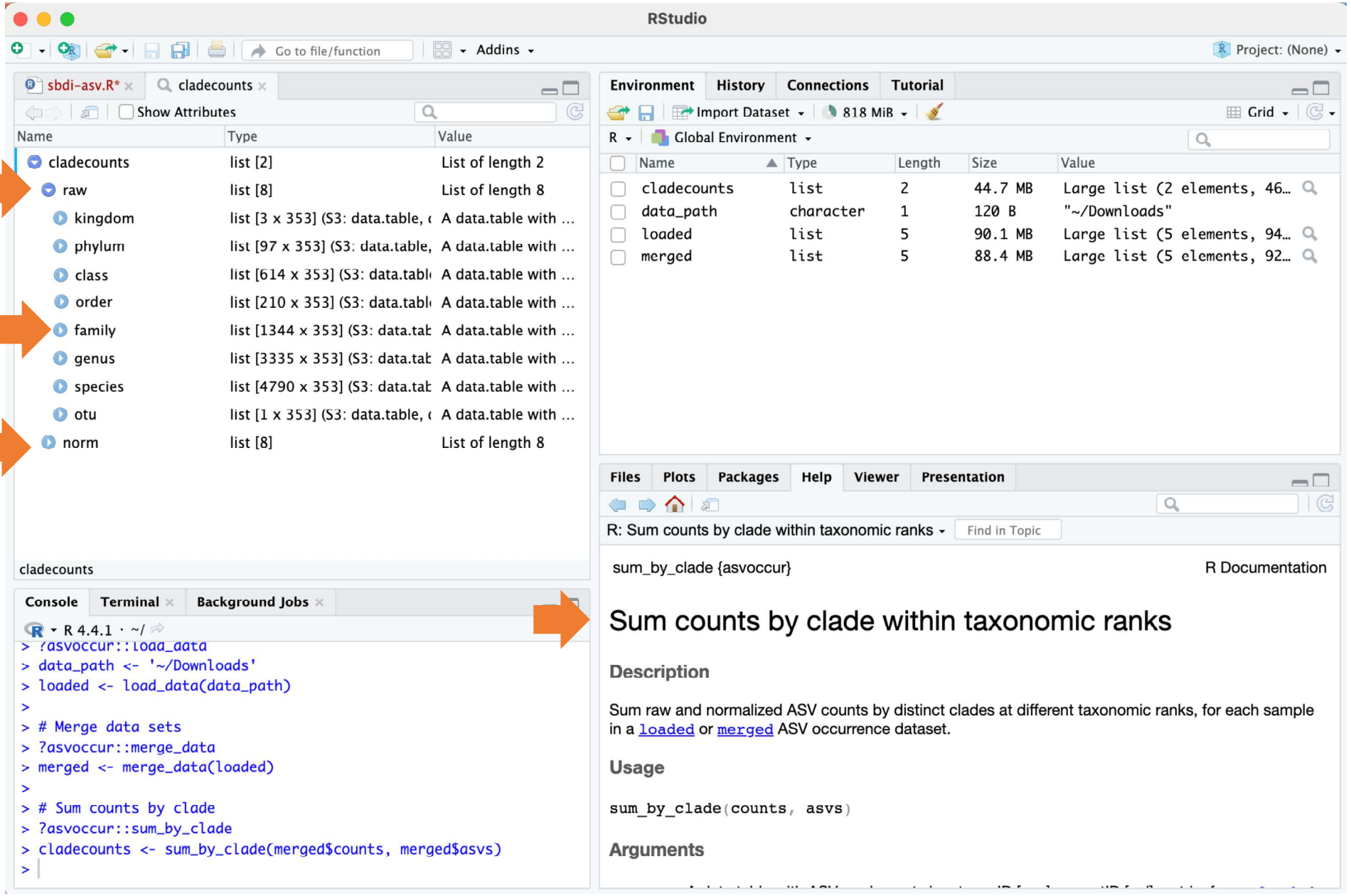

Start | ASV portal
asv-portal.biodiversitydata.se/index
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOGIN
Swedish ASV portal
Welcome to the portal of Swedish Amplicon Sequence Variants (ASVs) - an interface to sequence-based observations in SBDI
Search for ASVs and Bioatlas records using Basic Local Alignment Search Tool (BLAST)
BLAST
Search for ASVs and Bioatlas records using filters on sequencing details and taxonomy
FILTER
Submit your metabarcoding dataset to the ASV database and SBDI Bioatlas
SUBMIT
Download ASV occurrence datasets, in a condensed Darwin Core-like format
DOWNLOAD
WHO WE ARE
WHAT WE DO
WHERE WE ARE
INFORMATION

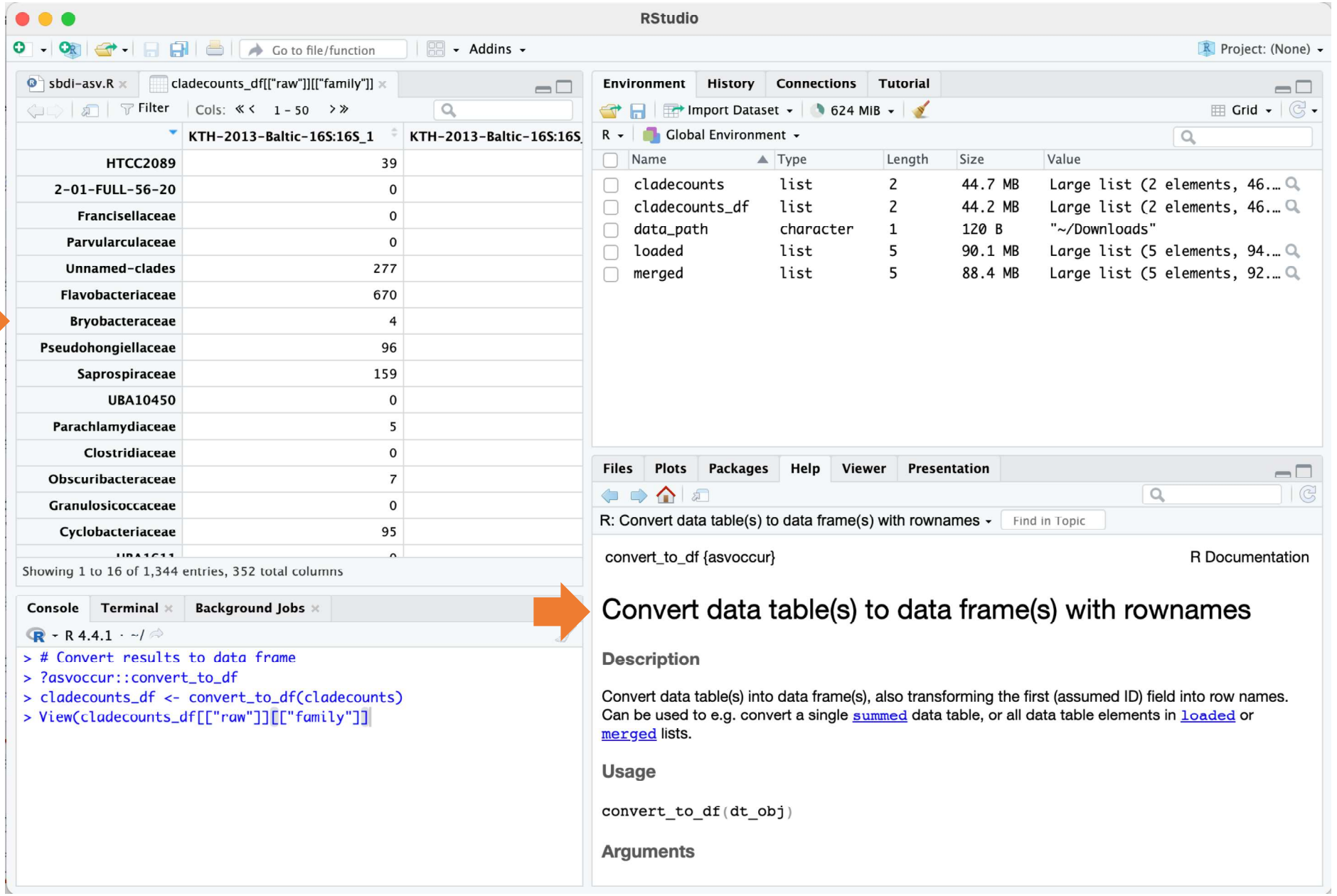

Download data | ASV portal
asv-portal.biodiversitydata.se/download
SBDI
Swedish Biodiversity Data Infrastructure
SBDI Home
Support form
SWEDISH ASV PORTAL
BLAST SEARCH
FILTER SEARCH
SUBMIT DATA
DOWNLOAD DATA
ABOUT
LOG OUT MARIA
Download data
Select and download complete ASV occurrence datasets in a modified Darwin Core (DwC) format. These files contain the same data as corresponding DwC archives served from the GBIF-Sweden Integrated Publishing Toolkit (IPT) website, but in a more condensed format. You can also inspect or download individual datasets via direct links, or retrieve lists of datasets in Excel or CSV format. You may have to actively allow downloads from our site in your browser preferences. Downloaded archives can be unpacked, merged and processed into ASV table format using the asvoccur R package.
Please contact SBDI support if you have any questions or suggestions.
Show 10 entries
Search: 16S
Target gene
Institution
Dataset name / IPT link
Download link
16S rRNA
KTH
16S: A comprehensive dataset on spatiotemporal variation of microbial plankton communities in the Baltic Sea
PRJEB55296-16S
16S rRNA
KTH
16S data from: Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea (Hu et al. 2016)
KTH-2013-Baltic-16S
EXCEL
CSV
Showing 1 to 2 of 2 entries (filtered from 10 total entries)
2 rows selected
Previous
1
Next
DOWNLOAD


Download data | ASV portal
biodiversitydata-se/asvoccur
github.com/biodiversitydata-se/asvoccur
README
License
MIT license
asvoccur
R tools for ASV occurrence data in SBDI.
Overview
The asvoccur R package, currently under development, provides tools for unpacking and processing ASV occurrence data and metadata downloaded from the Swedish ASV portal. It enables users to convert condensed DwC archives into ASV table format for easier downstream analysis in R, by using functions that load, merge, and aggregate ASV counts across taxonomic ranks.
Install
install.packages('devtools')
library(devtools)
install_github("biodiversitydata-se/asvoccur")
# or:
# install_github("biodiversitydata-se/asvoccur@develop")
library(asvoccur)
Run
Languages
R 100.0%


RStudio
Go to file/function
Addins
Project: (None)
sbdi-asv.R
loaded
Show Attributes
Name
Type
Value
loaded
list [5]
List of length 5
counts
list [2]
List of length 2
asvs
list [2]
List of length 2
KTH-2013-Baltic-...
list [3082 x 19] (S3: data.tabl
A data.table with ...
PRJEB55296-16S
list [40225 x 19] (S3: data.tab
A data.table with ...
events
list [2]
List of length 2
KTH-2013-Baltic-...
list [21 x 40] (S3: data.table, c
A data.table with ...
PRJEB55296-16S
list [331 x 40] (S3: data.table,
A data.table with ...
datasets
list [2]
List of length 2
emof
list [2]
List of length 2
loaded
Environment
History
Connections
Tutorial
Import Dataset
694 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
data_path
character
1
120 B
"~/Downloads"
loaded
list
5
90.1 MB
Large list (5 elements, 94...
Files
Plots
Packages
Help
Viewer
Presentation
R: Load downloaded ASV occurrence data
Find in Topic
load_data {asvoccur}
R Documentation
Load downloaded ASV occurrence data
Description
Load Amplicon Sequence Variant (ASV) occurrence data from 'Darwin Core (DwC)-like' archives downloaded from the Swedish ASP portal, https://asv-portal.biodiversitydata.se/.
Usage
load_data(data_path = './datasets');
Arguments
Console
Terminal
Background Jobs
R 4.4.1 · ~/
> lsf.str("package:asvoccur")
convert_to_df : function (dt_obj)
load_data : function (data_path = "./datasets")
merge_data : function (loaded, ds = NULL)
sum_by_clade : function (counts, asvs)
>
> # Load data
> ?asvoccur::load_data
> data_path <- '~/Downloads'
> loaded <- load_data(data_path)
> View(loaded)


RStudio
Go to file/function
Addins
Project: (None)
sbdi-asv.R
merged
loaded
Show Attributes
Name
Type
Value
merged
list [5]
List of length 5
counts
list [41355 x 353] (S3: data.ta
A data.table with ...
events
list [352 x 40] (S3: data.table,
A data.table with ...
datasets
list [2 x 2] (S3: data.table, dat
A data.table with ...
emof
list [352 x 65] (S3: data.table,
A data.table with ...
asvs
list [41355 x 15] (S3: data.tab
A data.table with ...
merged
Environment
History
Connections
Tutorial
Import Dataset
568 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
data_path
character
1
120 B
"~/Downloads"
loaded
list
5
90.1 MB
Large list (5 elements, 94...
merged
list
5
88.4 MB
Large list (5 elements, 92...
Files
Plots
Packages
Help
Viewer
Presentation
R: Merge data from different ASV occurrence datasets
Find in Topic
merge_data {asvoccur}
R Documentation
Merge data from different ASV occurrence datasets
Description
Merge data from different datasets previously loaded with load_data() function.
Usage
merge_data(loaded, ds = NULL)
Arguments
Console
Terminal
Background Jobs
R 4.4.1 · ~/
> # Load data
> ?asvoccur::load_data
> data_path <- '~/Downloads'
> loaded <- load_data(data_path)
> View(loaded)
> # Merge data
> ?asvoccur::merge_data
> merged <- merge_data(loaded)
> View(merged)
>

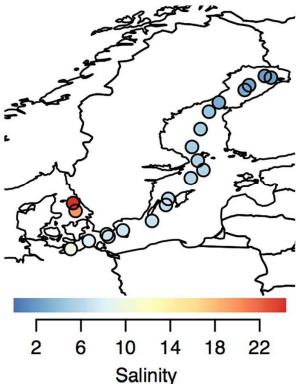

RStudio
Go to file/function
Addins
Project: (None)
sbdi-asv.R*
cladecounts
Show Attributes
Name
Type
Value
cladecounts
list [2]
List of length 2
raw
list [8]
List of length 8
kingdom
list [3 x 353] (S3: data.table,
A data.table with ...
phylum
list [97 x 353] (S3: data.table,
A data.table with ...
class
list [614 x 353] (S3: data.tabl
A data.table with ...
order
list [210 x 353] (S3: data.tabl
A data.table with ...
family
list [1344 x 353] (S3: data.tab
A data.table with ...
genus
list [3335 x 353] (S3: data.tab
A data.table with ...
species
list [4790 x 353] (S3: data.tab
A data.table with ...
otu
list [1 x 353] (S3: data.table,
A data.table with ...
norm
list [8]
List of length 8
cladecounts
Console
Terminal
Background Jobs
R 4.4.1 · ~/
> ?asvoccur::load_data
> data_path <- '~/Downloads'
> loaded <- load_data(data_path)
>
> # Merge data sets
> ?asvoccur::merge_data
> merged <- merge_data(loaded)
>
> # Sum counts by clade
> ?asvoccur::sum_by_clade
> cladecounts <- sum_by_clade(merged$counts, merged$asvs)
>
Environment
History
Connections
Tutorial
Import Dataset
818 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
cladecounts
list
2
44.7 MB
Large list (2 elements, 46...
data_path
character
1
120 B
"~/Downloads"
loaded
list
5
90.1 MB
Large list (5 elements, 94...
merged
list
5
88.4 MB
Large list (5 elements, 92...
Files
Plots
Packages
Help
Viewer
Presentation
R: Sum counts by clade within taxonomic ranks
Find in Topic
sum_by_clade {asvoccur}
R Documentation
Sum counts by clade within taxonomic ranks
Description
Sum raw and normalized ASV counts by distinct clades at different taxonomic ranks, for each sample in a loaded or merged ASV occurrence dataset.
Usage
sum_by_clade(counts, asvs)
Arguments


RStudio
Go to file/function
Addins
Project: (None)
sbdi-asv.R
cladecounts_df[["raw"]][["family"]]
Filter
Cols: « < 1 - 50 > »
KTH-2013-Baltic-16S:16S_1
KTH-2013-Baltic-16S:16S_
HTCC2089 39
2-01-FULL-56-20 0
Francisellaceae 0
Parvularculaceae 0
Unnamed-clades 277
Flavobacteriaceae 670
Bryobacteraceae 4
Pseudohongiellaceae 96
Saprospiraceae 159
UBA10450 0
Parachlamydiaceae 5
Clostridiaceae 0
Obscuribacteraceae 7
Granulosicoccaceae 0
Cyclobacteriaceae 95
Showing 1 to 16 of 1,344 entries, 352 total columns
Console
Terminal
Background Jobs
R 4.4.1 · ~/
> # Convert results to data frame
> ?asvoccur::convert_to_df
> cladecounts_df <- convert_to_df(cladecounts)
> View(cladecounts_df[["raw"]][["family"]])
Environment
History
Connections
Tutorial
Import Dataset
624 MiB
Grid
R
Global Environment
Name Type Length Size Value
cladecounts list 2 44.7 MB Large list (2 elements, 46....
cladecounts_df list 2 44.2 MB Large list (2 elements, 46....
data_path character 1 120 B "~/Downloads"
loaded list 5 90.1 MB Large list (5 elements, 94....
merged list 5 88.4 MB Large list (5 elements, 92....
Files
Plots
Packages
Help
Viewer
Presentation
R: Convert data table(s) to data frame(s) with rownames
Find in Topic
convert_to_df {asvoccur}
R Documentation
Convert data table(s) to data frame(s) with rownames
Description
Convert data table(s) into data frame(s), also transforming the first (assumed ID) field into row names. Can be used to e.g. convert a single summed data table, or all data table elements in loaded or merged lists.
Usage
convert_to_df(dt_obj)
Arguments

Diversity of Pico- to Mesoplankton along the 2000 km Salinity Gradient of the Baltic Sea

*Yue O. O. Hu[1], Bengt Karlson[2], Sophie Charvet[3] and Anders F. Andersson[1]**

[1] Science for Life Laboratory, Division of Gene Technology, School of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden, [2] Oceanography, Research & Development, Swedish Meteorological and Hydrological Institute, Gothenburg, Sweden, [3] Leibniz Institute for Baltic Sea Research Warnemünde, Rostock, Germany



21 surface water samples

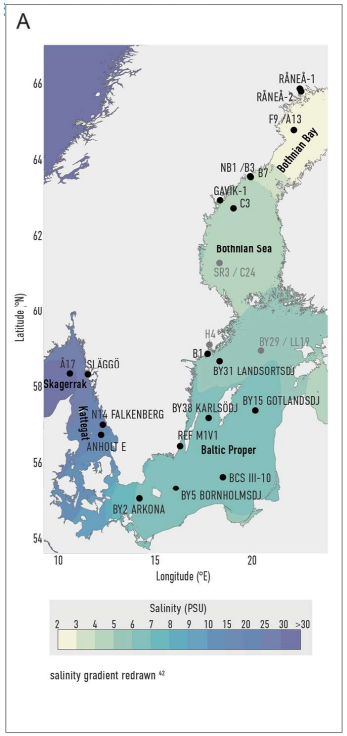
Sampled July 2013

OPEN

DATA DESCRIPTOR

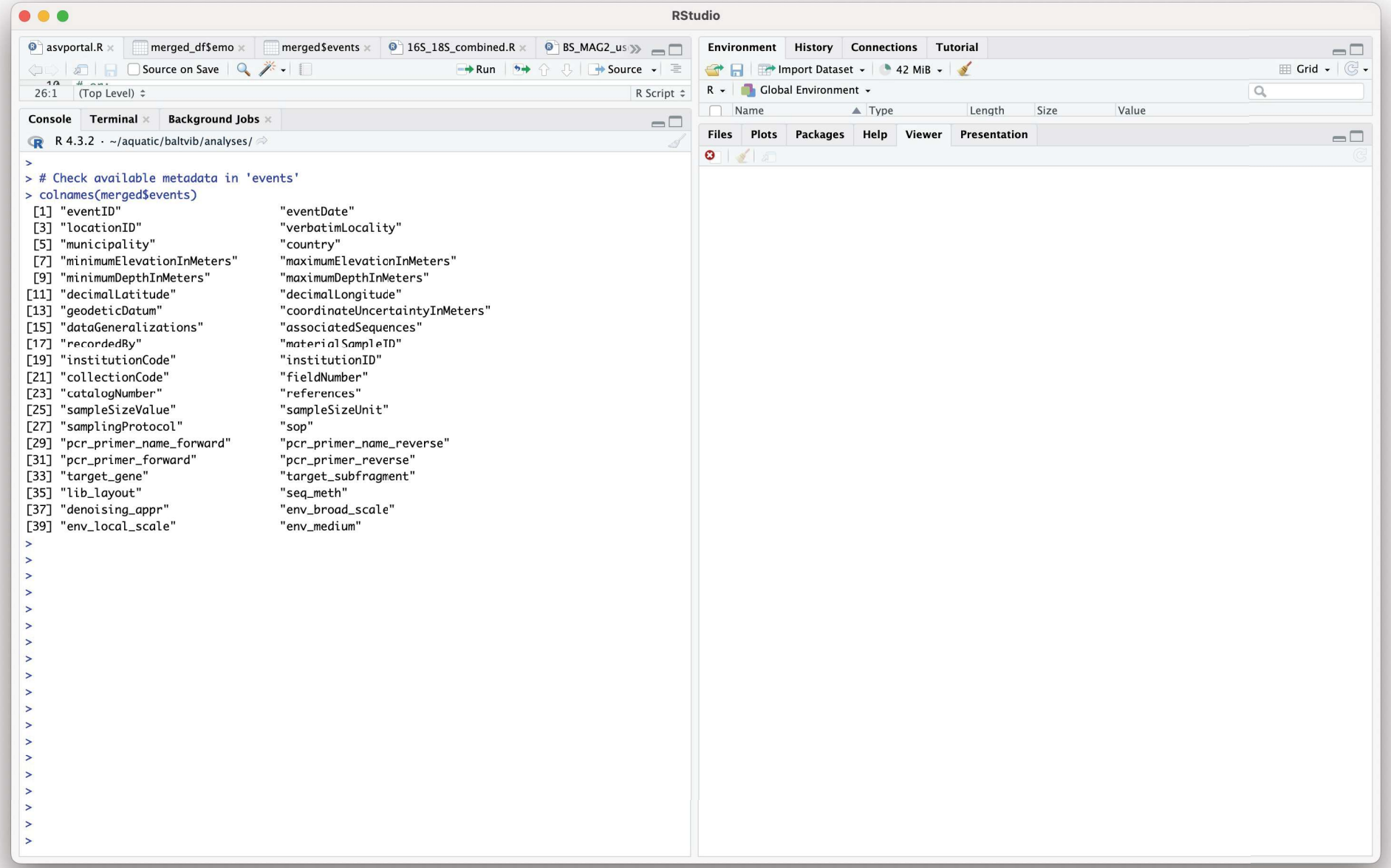
A comprehensive dataset on spatiotemporal variation of microbial plankton communities in the Baltic Sea

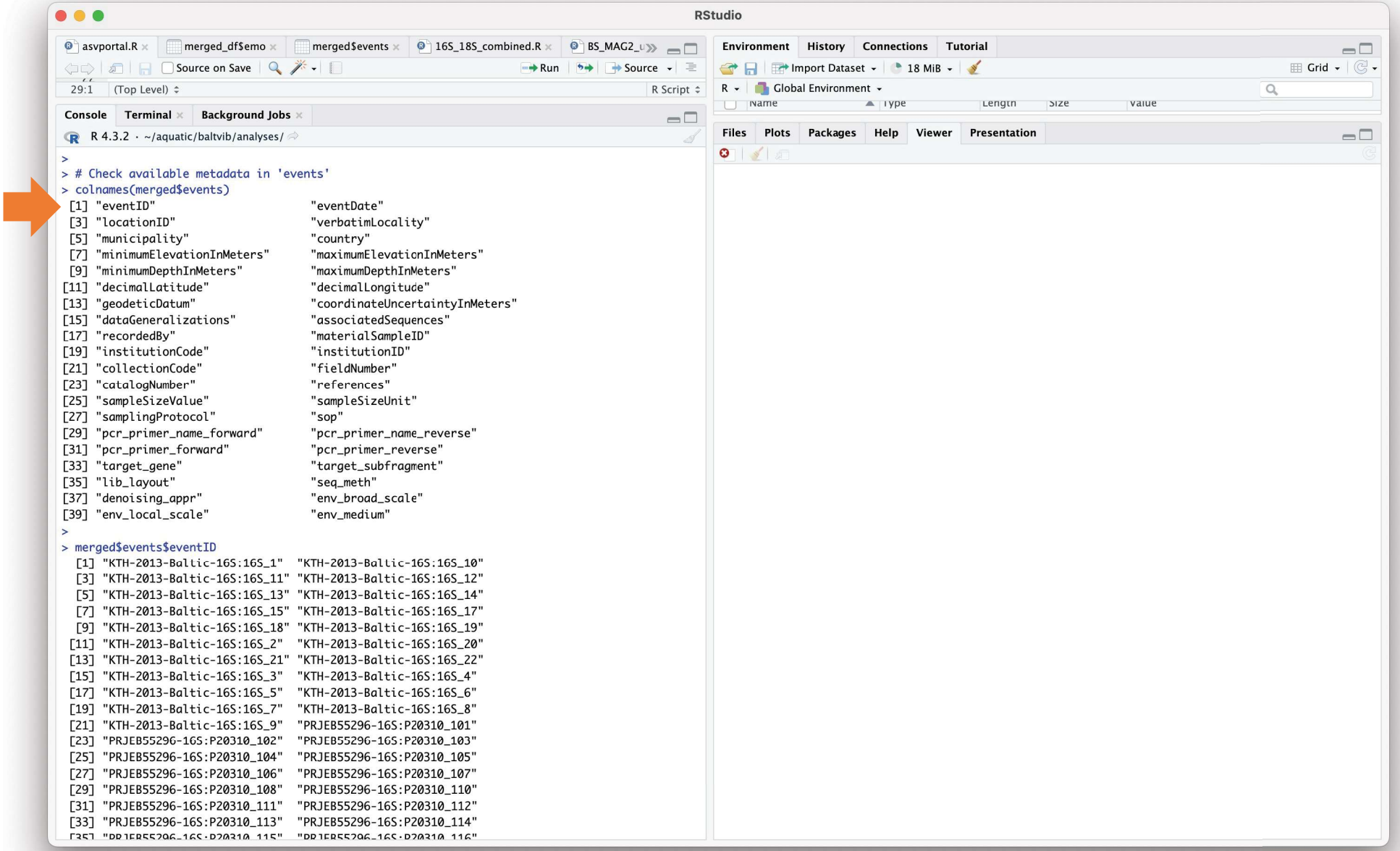
Meike A. C. Latz[1,2]✉, Agneta Andersson[3,4], Sonia Brugel[3,4], Mikael Hedblom[5], Krzysztof T. Jurdzinski[1], Bengt Karlson[6], Markus Lindh[5], Jenny Lycken[5], Anders Torstensson[5] & Anders F. Andersson[1]✉

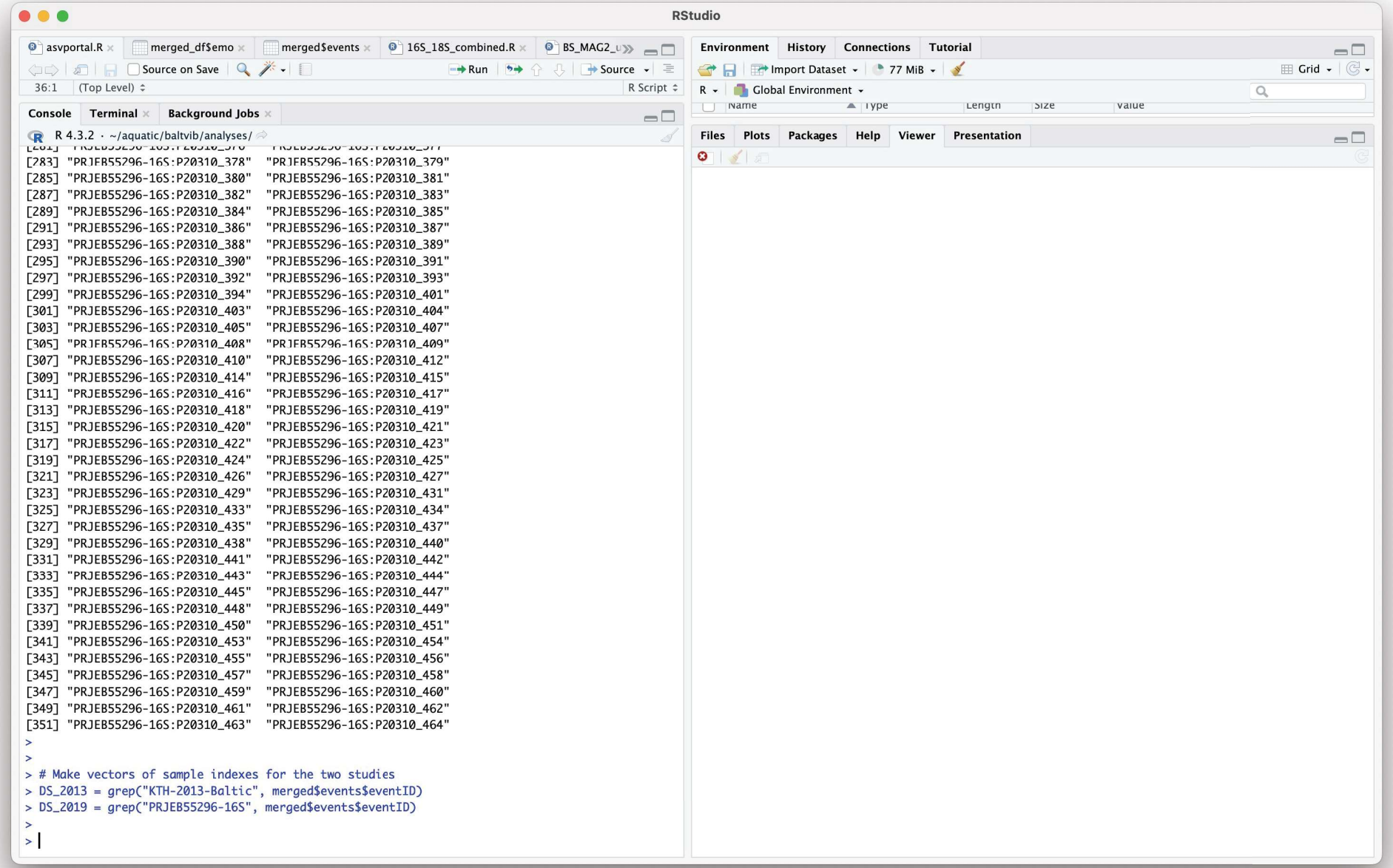


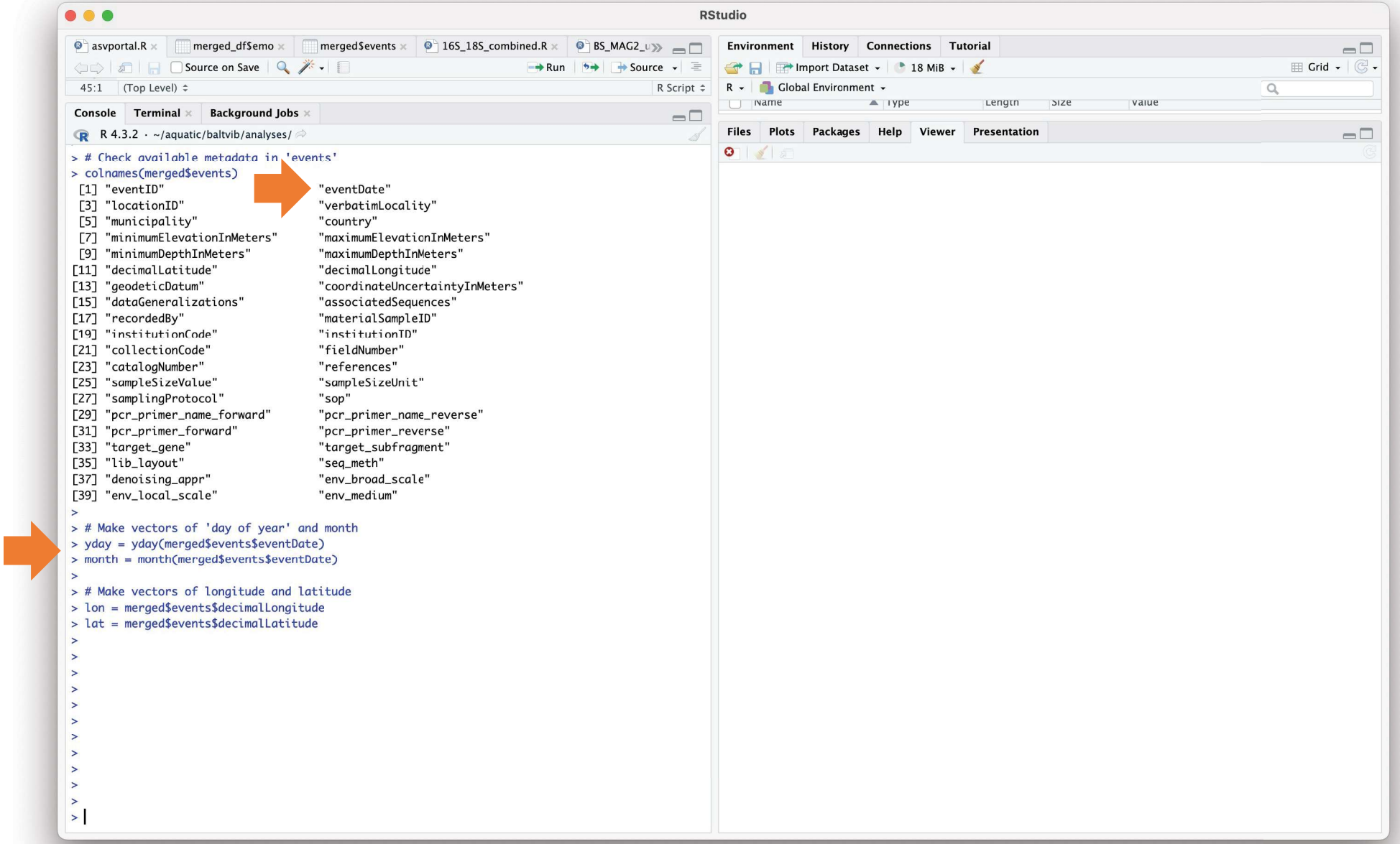
341 surface water samples

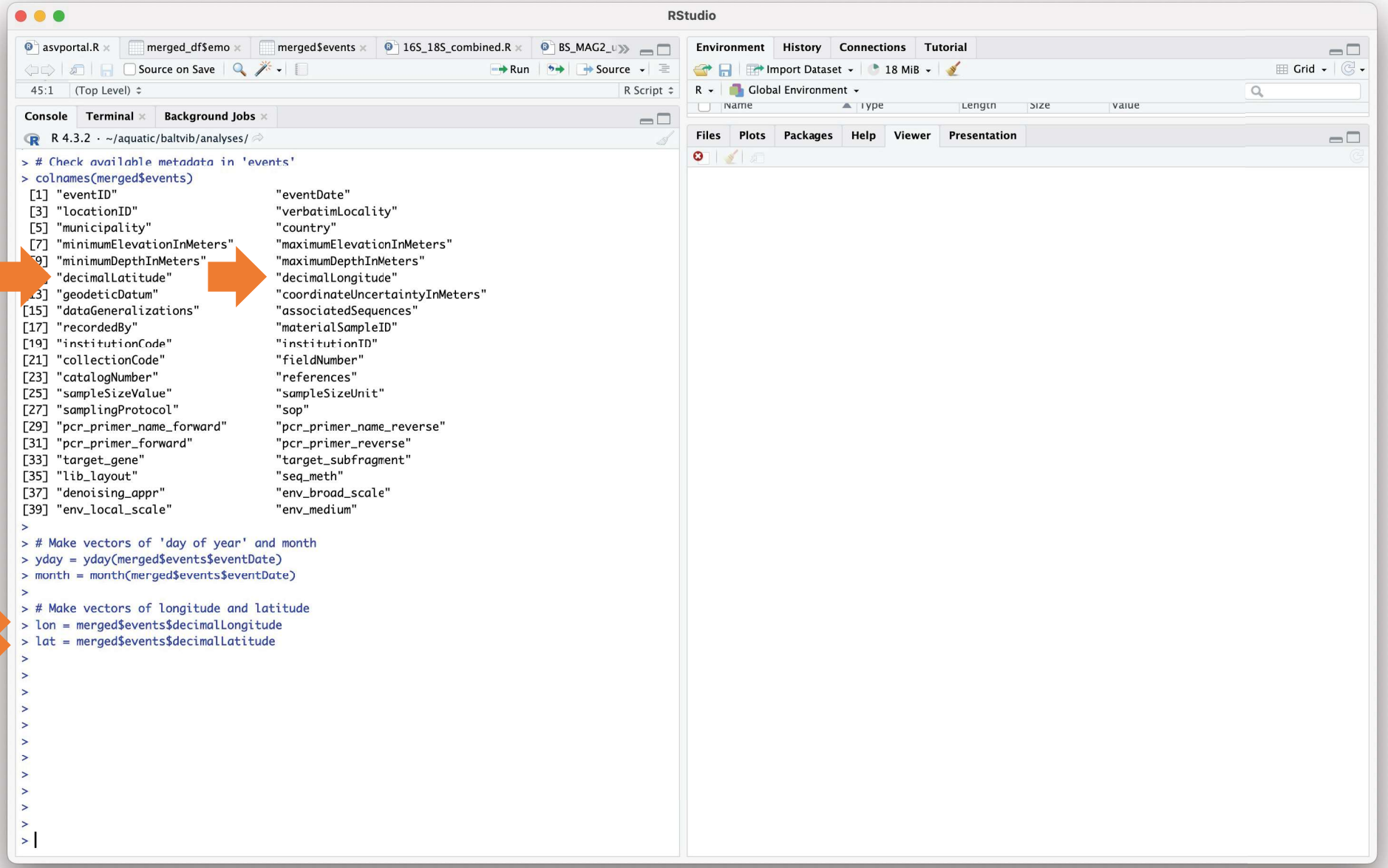
Sampled Jan 2019 – Feb 2020

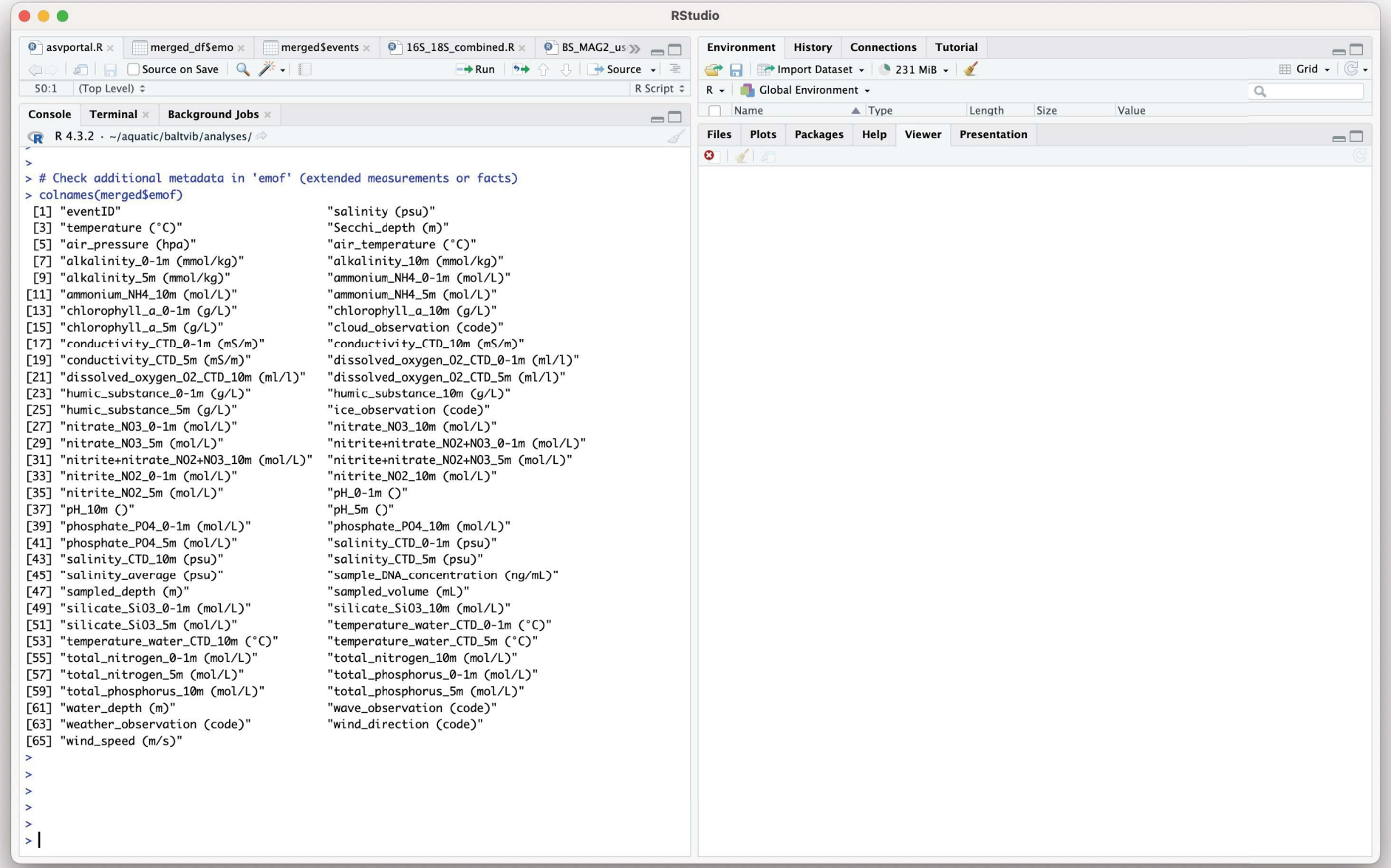

RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_us
Source on Save
Run
Source
26:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Check available metadata in 'events'
> colnames(merged$events)
[1] "eventID" "eventDate"
[3] "locationID" "verbatimLocality"
[5] "municipality" "country"
[7] "minimumElevationInMeters" "maximumElevationInMeters"
[9] "minimumDepthInMeters" "maximumDepthInMeters"
[11] "decimalLatitude" "decimalLongitude"
[13] "geodeticDatum" "coordinateUncertaintyInMeters"
[15] "dataGeneralizations" "associatedSequences"
[17] "recordedBy" "materialSampleID"
[19] "institutionCode" "institutionID"
[21] "collectionCode" "fieldNumber"
[23] "catalogNumber" "references"
[25] "sampleSizeValue" "sampleSizeUnit"
[27] "samplingProtocol" "sop"
[29] "pcr_primer_name_forward" "pcr_primer_name_reverse"
[31] "pcr_primer_forward" "pcr_primer_reverse"
[33] "target_gene" "target_subfragment"
[35] "lib_layout" "seq_meth"
[37] "denoising_appr" "env_broad_scale"
[39] "env_local_scale" "env_medium"
Environment
History
Connections
Tutorial
Import Dataset
42 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation

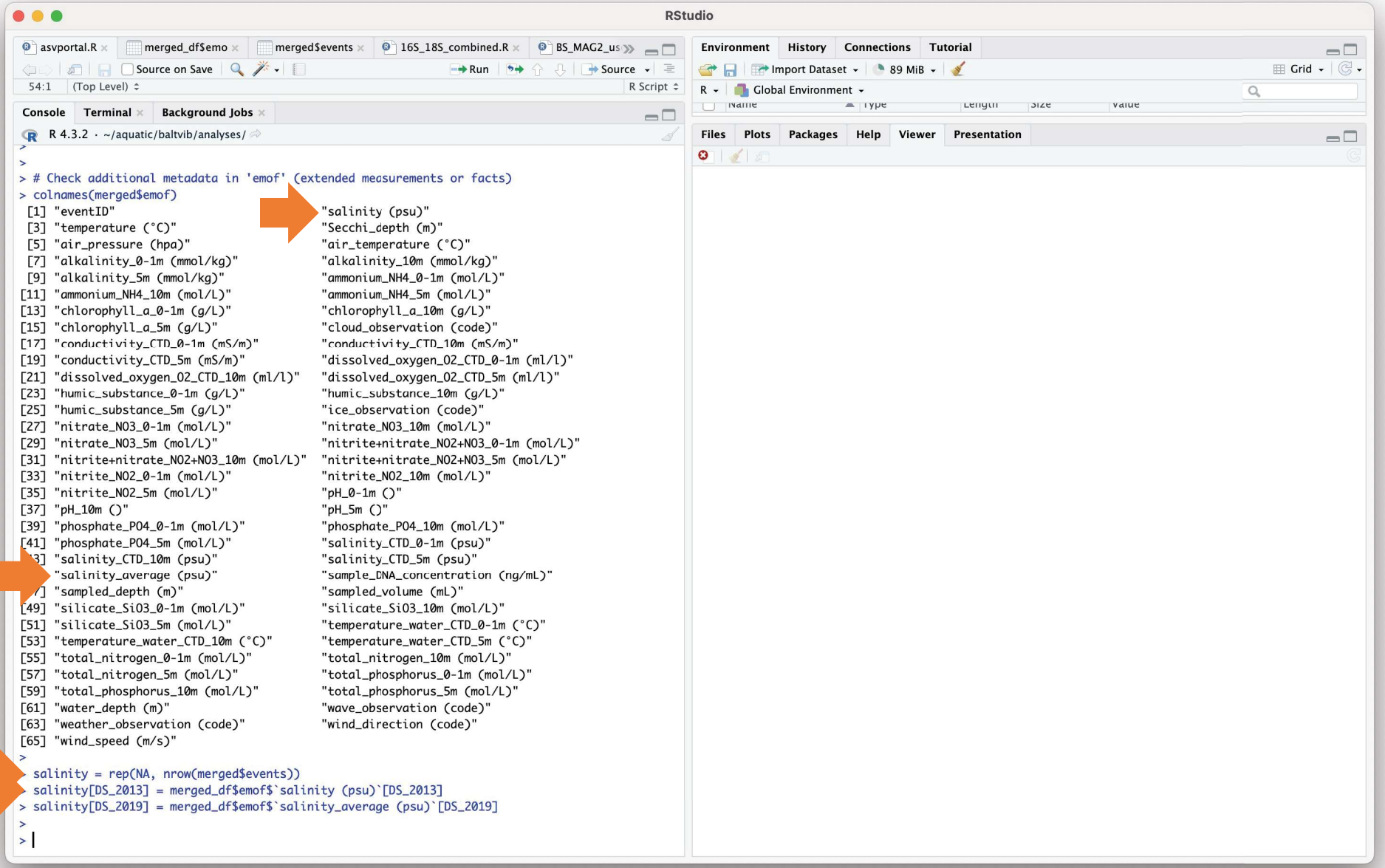

RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_u
Source on Save
Run
Source
29:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Check available metadata in 'events'
> colnames(merged$events)
[1] "eventID" "eventDate"
[3] "locationID" "verbatimLocality"
[5] "municipality" "country"
[7] "minimumElevationInMeters" "maximumElevationInMeters"
[9] "minimumDepthInMeters" "maximumDepthInMeters"
[11] "decimalLatitude" "decimalLongitude"
[13] "geodeticDatum" "coordinateUncertaintyInMeters"
[15] "dataGeneralizations" "associatedSequences"
[17] "recordedBy" "materialSampleID"
[19] "institutionCode" "institutionID"
[21] "collectionCode" "fieldNumber"
[23] "catalogNumber" "references"
[25] "sampleSizeValue" "sampleSizeUnit"
[27] "samplingProtocol" "sop"
[29] "pcr_primer_name_forward" "pcr_primer_name_reverse"
[31] "pcr_primer_forward" "pcr_primer_reverse"
[33] "target_gene" "target_subfragment"
[35] "lib_layout" "seq_meth"
[37] "denoising_appr" "env_broad_scale"
[39] "env_local_scale" "env_medium"
> merged$events$eventID
[1] "KTH-2013-Baltic-16S:16S_1" "KTH-2013-Baltic-16S:16S_10"
[3] "KTH-2013-Baltic-16S:16S_11" "KTH-2013-Baltic-16S:16S_12"
[5] "KTH-2013-Baltic-16S:16S_13" "KTH-2013-Baltic-16S:16S_14"
[7] "KTH-2013-Baltic-16S:16S_15" "KTH-2013-Baltic-16S:16S_17"
[9] "KTH-2013-Baltic-16S:16S_18" "KTH-2013-Baltic-16S:16S_19"
[11] "KTH-2013-Baltic-16S:16S_2" "KTH-2013-Baltic-16S:16S_20"
[13] "KTH-2013-Baltic-16S:16S_21" "KTH-2013-Baltic-16S:16S_22"
[15] "KTH-2013-Baltic-16S:16S_3" "KTH-2013-Baltic-16S:16S_4"
[17] "KTH-2013-Baltic-16S:16S_5" "KTH-2013-Baltic-16S:16S_6"
[19] "KTH-2013-Baltic-16S:16S_7" "KTH-2013-Baltic-16S:16S_8"
[21] "KTH-2013-Baltic-16S:16S_9" "PRJEB55296-16S:P20310_101"
[23] "PRJEB55296-16S:P20310_102" "PRJEB55296-16S:P20310_103"
[25] "PRJEB55296-16S:P20310_104" "PRJEB55296-16S:P20310_105"
[27] "PRJEB55296-16S:P20310_106" "PRJEB55296-16S:P20310_107"
[29] "PRJEB55296-16S:P20310_108" "PRJEB55296-16S:P20310_110"
[31] "PRJEB55296-16S:P20310_111" "PRJEB55296-16S:P20310_112"
[33] "PRJEB55296-16S:P20310_113" "PRJEB55296-16S:P20310_114"
Environment
History
Connections
Tutorial
Import Dataset
18 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation

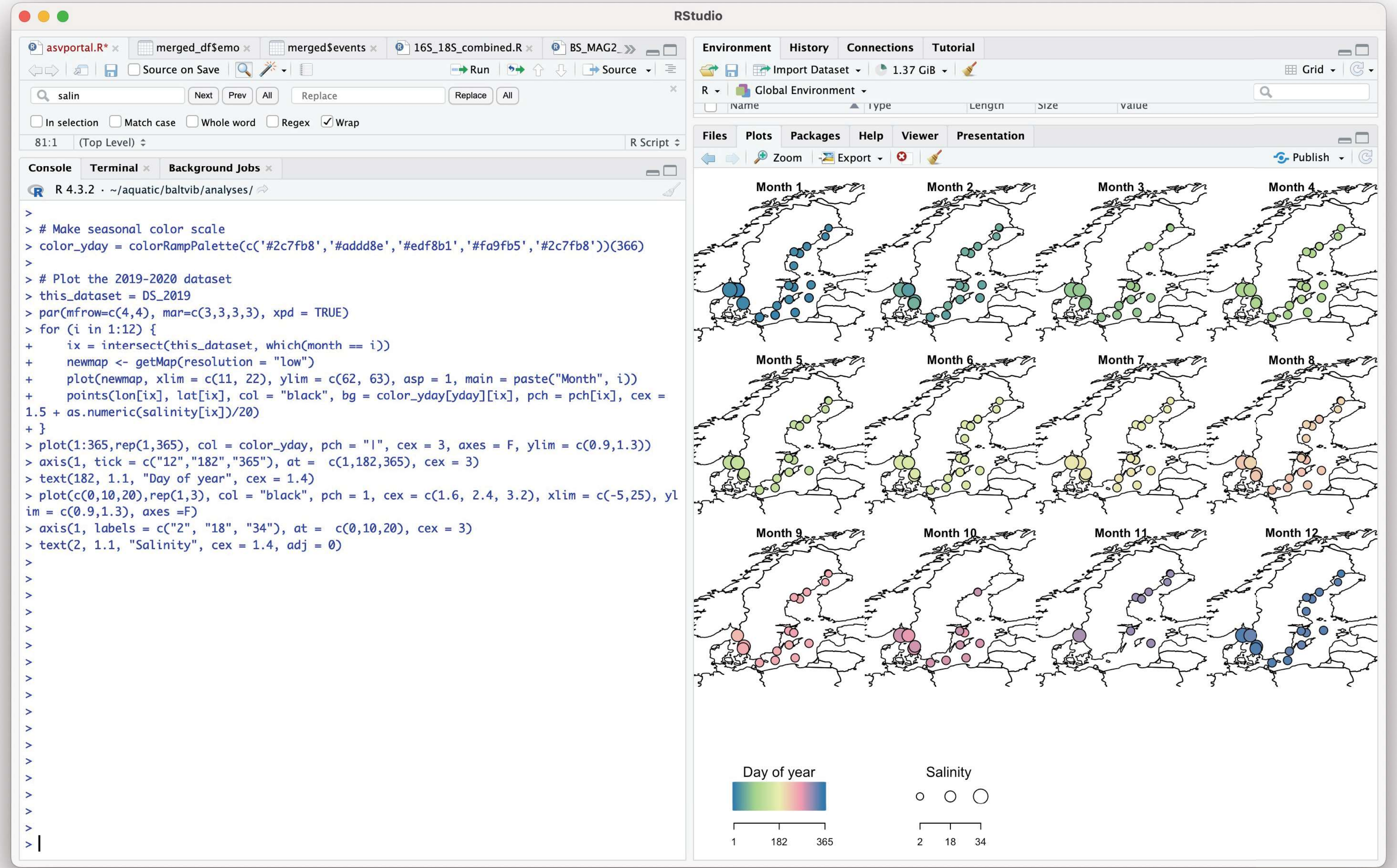

RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_u
Source on Save
Run
Source
36:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
[283] "PRJEB55296-16S:P20310_378" "PRJEB55296-16S:P20310_379"
[285] "PRJEB55296-16S:P20310_380" "PRJEB55296-16S:P20310_381"
[287] "PRJEB55296-16S:P20310_382" "PRJEB55296-16S:P20310_383"
[289] "PRJEB55296-16S:P20310_384" "PRJEB55296-16S:P20310_385"
[291] "PRJEB55296-16S:P20310_386" "PRJEB55296-16S:P20310_387"
[293] "PRJEB55296-16S:P20310_388" "PRJEB55296-16S:P20310_389"
[295] "PRJEB55296-16S:P20310_390" "PRJEB55296-16S:P20310_391"
[297] "PRJEB55296-16S:P20310_392" "PRJEB55296-16S:P20310_393"
[299] "PRJEB55296-16S:P20310_394" "PRJEB55296-16S:P20310_401"
[301] "PRJEB55296-16S:P20310_403" "PRJEB55296-16S:P20310_404"
[303] "PRJEB55296-16S:P20310_405" "PRJEB55296-16S:P20310_407"
[307] "PRJEB55296-16S:P20310_410" "PRJEB55296-16S:P20310_412"
[309] "PRJEB55296-16S:P20310_414" "PRJEB55296-16S:P20310_415"
[311] "PRJEB55296-16S:P20310_416" "PRJEB55296-16S:P20310_417"
[313] "PRJEB55296-16S:P20310_418" "PRJEB55296-16S:P20310_419"
[315] "PRJEB55296-16S:P20310_420" "PRJEB55296-16S:P20310_421"
[317] "PRJEB55296-16S:P20310_422" "PRJEB55296-16S:P20310_423"
[319] "PRJEB55296-16S:P20310_424" "PRJEB55296-16S:P20310_425"
[321] "PRJEB55296-16S:P20310_426" "PRJEB55296-16S:P20310_427"
[323] "PRJEB55296-16S:P20310_429" "PRJEB55296-16S:P20310_431"
[325] "PRJEB55296-16S:P20310_433" "PRJEB55296-16S:P20310_434"
[327] "PRJEB55296-16S:P20310_435" "PRJEB55296-16S:P20310_437"
[329] "PRJEB55296-16S:P20310_438" "PRJEB55296-16S:P20310_440"
[331] "PRJEB55296-16S:P20310_441" "PRJEB55296-16S:P20310_442"
[333] "PRJEB55296-16S:P20310_443" "PRJEB55296-16S:P20310_444"
[335] "PRJEB55296-16S:P20310_445" "PRJEB55296-16S:P20310_447"
[337] "PRJEB55296-16S:P20310_448" "PRJEB55296-16S:P20310_449"
[339] "PRJEB55296-16S:P20310_450" "PRJEB55296-16S:P20310_451"
[341] "PRJEB55296-16S:P20310_453" "PRJEB55296-16S:P20310_454"
[343] "PRJEB55296-16S:P20310_455" "PRJEB55296-16S:P20310_456"
[345] "PRJEB55296-16S:P20310_457" "PRJEB55296-16S:P20310_458"
[347] "PRJEB55296-16S:P20310_459" "PRJEB55296-16S:P20310_460"
[349] "PRJEB55296-16S:P20310_461" "PRJEB55296-16S:P20310_462"
[351] "PRJEB55296-16S:P20310_463" "PRJEB55296-16S:P20310_464"
> # Make vectors of sample indexes for the two studies
> DS_2013 = grep("KTH-2013-Baltic", merged$events$eventID)
> DS_2019 = grep("PRJEB55296-16S", merged$events$eventID)
Environment
History
Connections
Tutorial
Import Dataset
77 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation

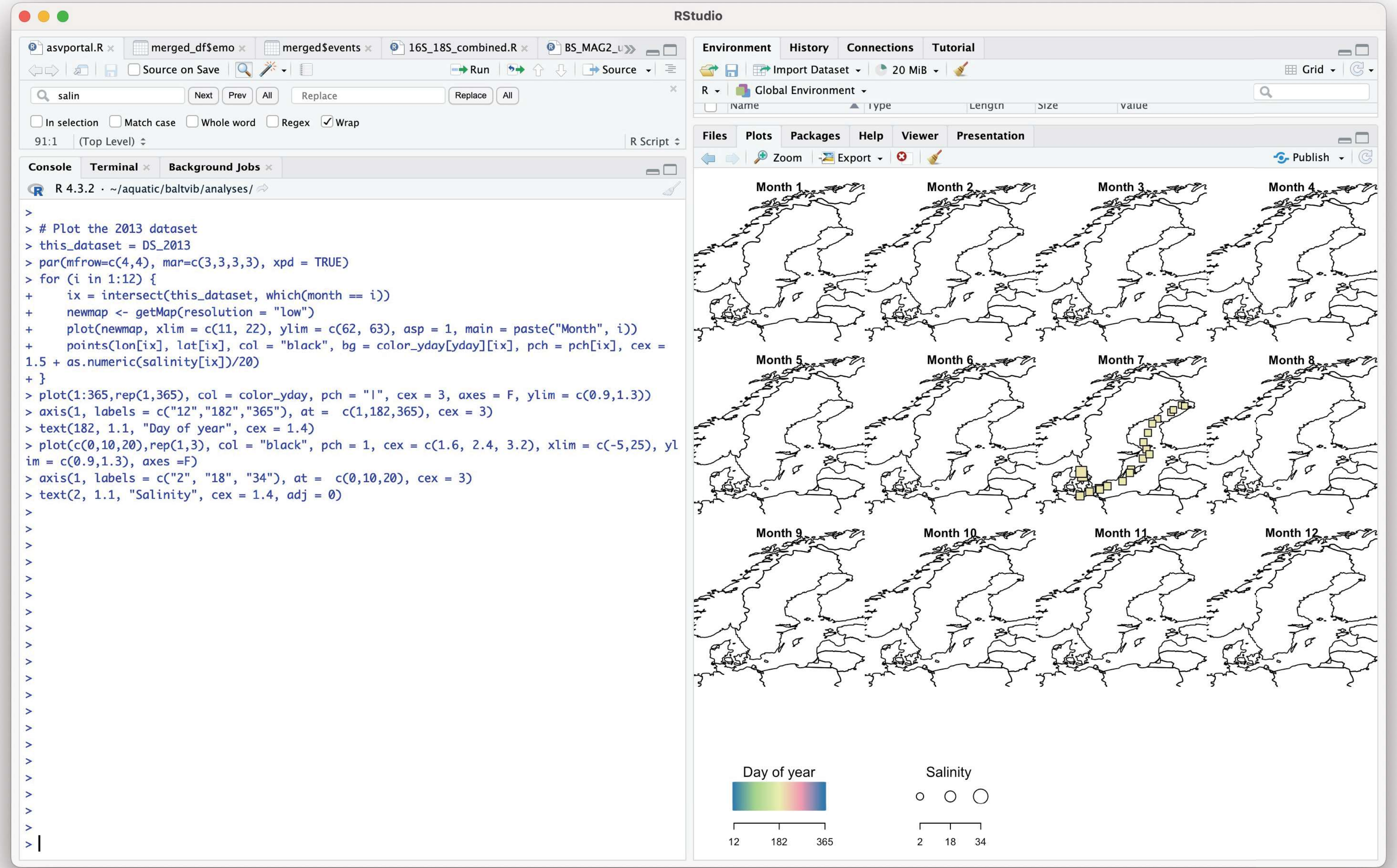

RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_u
Source on Save
Run
Source
45:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Check available metadata in 'events'
> colnames(merged$events)
[1] "eventID" "eventDate"
[3] "locationID" "verbatimLocality"
[5] "municipality" "country"
[7] "minimumElevationInMeters" "maximumElevationInMeters"
[9] "minimumDepthInMeters" "maximumDepthInMeters"
[11] "decimalLatitude" "decimalLongitude"
[13] "geodeticDatum" "coordinateUncertaintyInMeters"
[15] "dataGeneralizations" "associatedSequences"
[17] "recordedBy" "materialSampleID"
[19] "institutionCode" "institutionID"
[21] "collectionCode" "fieldNumber"
[23] "catalogNumber" "references"
[25] "sampleSizeValue" "sampleSizeUnit"
[27] "samplingProtocol" "sop"
[29] "pcr_primer_name_forward" "pcr_primer_name_reverse"
[31] "pcr_primer_forward" "pcr_primer_reverse"
[33] "target_gene" "target_subfragment"
[35] "lib_layout" "seq_meth"
[37] "denoising_appr" "env_broad_scale"
[39] "env_local_scale" "env_medium"
> # Make vectors of 'day of year' and month
> yday = yday(merged$events$eventDate)
> month = month(merged$events$eventDate)
> # Make vectors of longitude and latitude
> lon = merged$events$decimalLongitude
> lat = merged$events$decimalLatitude
Environment
History
Connections
Tutorial
Import Dataset
18 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation


RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_u
Source on Save
Run
Source
45:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Check available metadata in 'events'
> colnames(merged$events)
[1] "eventID" "eventDate"
[3] "locationID" "verbatimLocality"
[5] "municipality" "country"
[7] "minimumElevationInMeters" "maximumElevationInMeters"
[9] "minimumDepthInMeters" "maximumDepthInMeters"
"decimalLatitude" "decimalLongitude"
[13] "geodeticDatum" "coordinateUncertaintyInMeters"
[15] "dataGeneralizations" "associatedSequences"
[17] "recordedBy" "materialSampleID"
[19] "institutionCode" "institutionID"
[21] "collectionCode" "fieldNumber"
[23] "catalogNumber" "references"
[25] "sampleSizeValue" "sampleSizeUnit"
[27] "samplingProtocol" "sop"
[29] "pcr_primer_name_forward" "pcr_primer_name_reverse"
[31] "pcr_primer_forward" "pcr_primer_reverse"
[33] "target_gene" "target_subfragment"
[35] "lib_layout" "seq_meth"
[37] "denoising_appr" "env_broad_scale"
[39] "env_local_scale" "env_medium"
> # Make vectors of 'day of year' and month
> yday = yday(merged$events$eventDate)
> month = month(merged$events$eventDate)
> # Make vectors of longitude and latitude
> lon = merged$events$decimalLongitude
> lat = merged$events$decimalLatitude
Environment
History
Connections
Tutorial
Import Dataset
18 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation


RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_us
Source on Save
Run
Source
50:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Check additional metadata in 'emof' (extended measurements or facts)
> colnames(merged$emof)
[1] "eventID" "salinity (psu)"
[3] "temperature (°C)" "Secchi_depth (m)"
[5] "air_pressure (hpa)" "air_temperature (°C)"
[7] "alkalinity_0-1m (mmol/kg)" "alkalinity_10m (mmol/kg)"
[9] "alkalinity_5m (mmol/kg)" "ammonium_NH4_0-1m (mol/L)"
[11] "ammonium_NH4_10m (mol/L)" "ammonium_NH4_5m (mol/L)"
[13] "chlorophyll_a_0-1m (g/L)" "chlorophyll_a_10m (g/L)"
[15] "chlorophyll_a_5m (g/L)" "cloud_observation (code)"
[17] "conductivity_CTD_0-1m (mS/m)" "conductivity_CTD_10m (mS/m)"
[19] "conductivity_CTD_5m (mS/m)" "dissolved_oxygen_O2_CTD_0-1m (ml/l)"
[21] "dissolved_oxygen_O2_CTD_10m (ml/l)" "dissolved_oxygen_O2_CTD_5m (ml/l)"
[23] "humic_substance_0-1m (g/L)" "humic_substance_10m (g/L)"
[25] "humic_substance_5m (g/L)" "ice_observation (code)"
[27] "nitrate_NO3_0-1m (mol/L)" "nitrate_NO3_10m (mol/L)"
[29] "nitrate_NO3_5m (mol/L)" "nitrite+nitrate_NO2+NO3_0-1m (mol/L)"
[31] "nitrite+nitrate_NO2+NO3_10m (mol/L)" "nitrite+nitrate_NO2+NO3_5m (mol/L)"
[33] "nitrite_NO2_0-1m (mol/L)" "nitrite_NO2_10m (mol/L)"
[35] "nitrite_NO2_5m (mol/L)" "pH_0-1m ()"
[37] "pH_10m ()" "pH_5m ()"
[39] "phosphate_PO4_0-1m (mol/L)" "phosphate_PO4_10m (mol/L)"
[41] "phosphate_PO4_5m (mol/L)" "salinity_CTD_0-1m (psu)"
[43] "salinity_CTD_10m (psu)" "salinity_CTD_5m (psu)"
[45] "salinity_average (psu)" "sample_DNA_concentration (ng/mL)"
[47] "sampled_depth (m)" "sampled_volume (mL)"
[49] "silicate_SiO3_0-1m (mol/L)" "silicate_SiO3_10m (mol/L)"
[51] "silicate_SiO3_5m (mol/L)" "temperature_water_CTD_0-1m (°C)"
[53] "temperature_water_CTD_10m (°C)" "temperature_water_CTD_5m (°C)"
[55] "total_nitrogen_0-1m (mol/L)" "total_nitrogen_10m (mol/L)"
[57] "total_nitrogen_5m (mol/L)" "total_phosphorus_0-1m (mol/L)"
[59] "total_phosphorus_10m (mol/L)" "total_phosphorus_5m (mol/L)"
[61] "water_depth (m)" "wave_observation (code)"
[63] "weather_observation (code)" "wind_direction (code)"
[65] "wind_speed (m/s)"
Environment
History
Connections
Tutorial
Import Dataset
231 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation


RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_us
Source on Save
Run
Source
54:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Check additional metadata in 'emof' (extended measurements or facts)
> colnames(merged$emof)
[1] "eventID"
[2] "salinity (psu)"
[3] "temperature (°C)"
[4] "Secchi_depth (m)"
[5] "air_pressure (hpa)"
[6] "air_temperature (°C)"
[7] "alkalinity_0-1m (mmol/kg)"
[8] "alkalinity_10m (mmol/kg)"
[9] "alkalinity_5m (mmol/kg)"
[10] "ammonium_NH4_0-1m (mol/L)"
[11] "ammonium_NH4_10m (mol/L)"
[12] "ammonium_NH4_5m (mol/L)"
[13] "chlorophyll_a_0-1m (g/L)"
[14] "chlorophyll_a_10m (g/L)"
[15] "chlorophyll_a_5m (g/L)"
[16] "cloud_observation (code)"
[17] "conductivity_CTD_0-1m (mS/m)"
[18] "conductivity_CTD_10m (mS/m)"
[19] "conductivity_CTD_5m (mS/m)"
[20] "dissolved_oxygen_O2_CTD_0-1m (ml/l)"
[21] "dissolved_oxygen_O2_CTD_10m (ml/l)"
[22] "dissolved_oxygen_O2_CTD_5m (ml/l)"
[23] "humic_substance_0-1m (g/L)"
[24] "humic_substance_10m (g/L)"
[25] "humic_substance_5m (g/L)"
[26] "ice_observation (code)"
[27] "nitrate_NO3_0-1m (mol/L)"
[28] "nitrate_NO3_10m (mol/L)"
[29] "nitrate_NO3_5m (mol/L)"
[30] "nitrite+nitrate_NO2+NO3_0-1m (mol/L)"
[31] "nitrite+nitrate_NO2+NO3_10m (mol/L)"
[32] "nitrite+nitrate_NO2+NO3_5m (mol/L)"
[33] "nitrite_NO2_0-1m (mol/L)"
[34] "nitrite_NO2_10m (mol/L)"
[35] "nitrite_NO2_5m (mol/L)"
[36] "pH_0-1m ()"
[37] "pH_10m ()"
[38] "pH_5m ()"
[39] "phosphate_PO4_0-1m (mol/L)"
[40] "phosphate_PO4_10m (mol/L)"
[41] "phosphate_PO4_5m (mol/L)"
[42] "salinity_CTD_0-1m (psu)"
[43] "salinity_CTD_10m (psu)"
[44] "salinity_CTD_5m (psu)"
[45] "salinity_average (psu)"
[46] "sample_DNA_concentration (ng/mL)"
[47] "sampled_depth (m)"
[48] "sampled_volume (mL)"
[49] "silicate_SiO3_0-1m (mol/L)"
[50] "silicate_SiO3_10m (mol/L)"
[51] "silicate_SiO3_5m (mol/L)"
[52] "temperature_water_CTD_0-1m (°C)"
[53] "temperature_water_CTD_10m (°C)"
[54] "temperature_water_CTD_5m (°C)"
[55] "total_nitrogen_0-1m (mol/L)"
[56] "total_nitrogen_10m (mol/L)"
[57] "total_nitrogen_5m (mol/L)"
[58] "total_phosphorus_0-1m (mol/L)"
[59] "total_phosphorus_10m (mol/L)"
[60] "total_phosphorus_5m (mol/L)"
[61] "water_depth (m)"
[62] "wave_observation (code)"
[63] "weather_observation (code)"
[64] "wind_direction (code)"
[65] "wind_speed (m/s)"
> salinity = rep(NA, nrow(merged$events))
> salinity[DS_2013] = merged_df$emof$`salinity (psu)`[DS_2013]
> salinity[DS_2019] = merged_df$emof$`salinity_average (psu)`[DS_2019]
Environment
History
Connections
Tutorial
Import Dataset
89 MiB
Grid
R
Global Environment
Files
Plots
Packages
Help
Viewer
Presentation


RStudio
asvportal.R*
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_
Source on Save
Run
Source
salin
Next
Prev
All
Replace
In selection
Match case
Whole word
Regex
Wrap
81:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Make seasonal color scale
> color_yday = colorRampPalette(c('#2c7fb8','#addd8e','#edf8b1','#fa9fb5','#2c7fb8'))(366)
> # Plot the 2019-2020 dataset
> this_dataset = DS_2019
> par(mfrow=c(4,4), mar=c(3,3,3,3), xpd = TRUE)
> for (i in 1:12) {
+ ix = intersect(this_dataset, which(month == i))
+ newmap <- getMap(resolution = "low")
+ plot(newmap, xlim = c(11, 22), ylim = c(62, 63), asp = 1, main = paste("Month", i))
+ points(lon[ix], lat[ix], col = "black", bg = color_yday[yday][ix], pch = pch[ix], cex = 1.5 + as.numeric(salinity[ix])/20)
+ }
> plot(1:365,rep(1,365), col = color_yday, pch = "|", cex = 3, axes = F, ylim = c(0.9,1.3))
> axis(1, tick = c("12","182","365"), at = c(1,182,365), cex = 3)
> text(182, 1.1, "Day of year", cex = 1.4)
> plot(c(0,10,20),rep(1,3), col = "black", pch = 1, cex = c(1.6, 2.4, 3.2), xlim = c(-5,25), ylim = c(0.9,1.3), axes =F)
> axis(1, labels = c("2", "18", "34"), at = c(0,10,20), cex = 3)
> text(2, 1.1, "Salinity", cex = 1.4, adj = 0)
Environment
History
Connections
Tutorial
Import Dataset
1.37 GiB
Grid
Global Environment
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
Month 1
Month 2
Month 3
Month 4
Month 5
Month 6
Month 7
Month 8
Month 9
Month 10
Month 11
Month 12
Day of year
1
182
365
Salinity
2
18
34


RStudio
asvportal.R
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_u
Source on Save
Run
Source
salin
Next
Prev
All
Replace
Replace
All
In selection
Match case
Whole word
Regex
Wrap
91:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Plot the 2013 dataset
> this_dataset = DS_2013
> par(mfrow=c(4,4), mar=c(3,3,3,3), xpd = TRUE)
> for (i in 1:12) {
+ ix = intersect(this_dataset, which(month == i))
+ newmap <- getMap(resolution = "low")
+ plot(newmap, xlim = c(11, 22), ylim = c(62, 63), asp = 1, main = paste("Month", i))
+ points(lon[ix], lat[ix], col = "black", bg = color_yday[yday][ix], pch = pch[ix], cex = 1.5 + as.numeric(salinity[ix])/20)
+ }
> plot(1:365,rep(1,365), col = color_yday, pch = "|", cex = 3, axes = F, ylim = c(0.9,1.3))
> axis(1, labels = c("12","182","365"), at = c(1,182,365), cex = 3)
> text(182, 1.1, "Day of year", cex = 1.4)
> plot(c(0,10,20),rep(1,3), col = "black", pch = 1, cex = c(1.6, 2.4, 3.2), xlim = c(-5,25), ylim = c(0.9,1.3), axes =F)
> axis(1, labels = c("2", "18", "34"), at = c(0,10,20), cex = 3)
> text(2, 1.1, "Salinity", cex = 1.4, adj = 0)
Environment
History
Connections
Tutorial
Import Dataset
20 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
Month 1
Month 2
Month 3
Month 4
Month 5
Month 6
Month 7
Month 8
Month 9
Month 10
Month 11
Month 12
Day of year
12
182
365
Salinity
2
18
34

RStudio

asvportal.R | merged_df$emo | merged$events | 16S_18S_combined.R | BS_MAG2_used_in_p

Source on Save | Run | Source

salin | Next | Prev | All | Replace | Replace | All

In selection | Match case | Whole word | Regex | Wrap

128:1 | (Top Level) | R Script

Console | Terminal | Background Jobs

R 4.3.2 · ~/aquatic/baltvib/analyses/

```
> 
> # PCoA on Bray-Curtis distances
> pcoa <- pcoa(bray_dist, correction = "cailliez")
> 
> 
> # Plotting the color and size legends
> layout(matrix(c(1,2,3,3,3,3), 3, 2, byrow = T))
> par(mar=c(5,5,1,1), xpd = TRUE, cex.axis = 1.5)
> plot(1:365,rep(1,365), col = color_yday, pch = "|", cex = 3, axes = F, ylim = c(0.9,1.3), ylab = "",
xlab = "")
> axis(1, labels = c("12","182","365"), at =  c(1,182,365), cex = 3)
> text(182, 1.1, "Day of year", cex = 2)
> plot(c(0,10,20),rep(1,3), col = "black", pch = 1, cex = c(1.6, 2.4, 3.2), xlim = c(-5,25), ylim = c
(0.9,1.3), axes =F, ylab = "", xlab = "")
> axis(1, labels = c("2", "18", "34"), at =  c(0,10,20), cex = 3)
> text(10, 1.1, "Salinity", cex = 2)
> 
> 
> # Plotting the PCoA
> plot(pcoa$vectors[,1], pcoa$vectors[,2], pch = pch, col = "white", bg = "white", xlab = "PC1", ylab
= "PC2")
> # plotting the DS_2019 samples
> points(pcoa$vectors[DS_2019,1], pcoa$vectors[DS_2019,2], pch = pch[DS_2019], col = "black", bg = col
or_yday[yday[DS_2019]], cex = 1.0 + as.numeric(salinity[DS_2019])/10, xlab = xlab, ylab = ylab)
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
> 
```

Environment | History | Connections | Tutorial

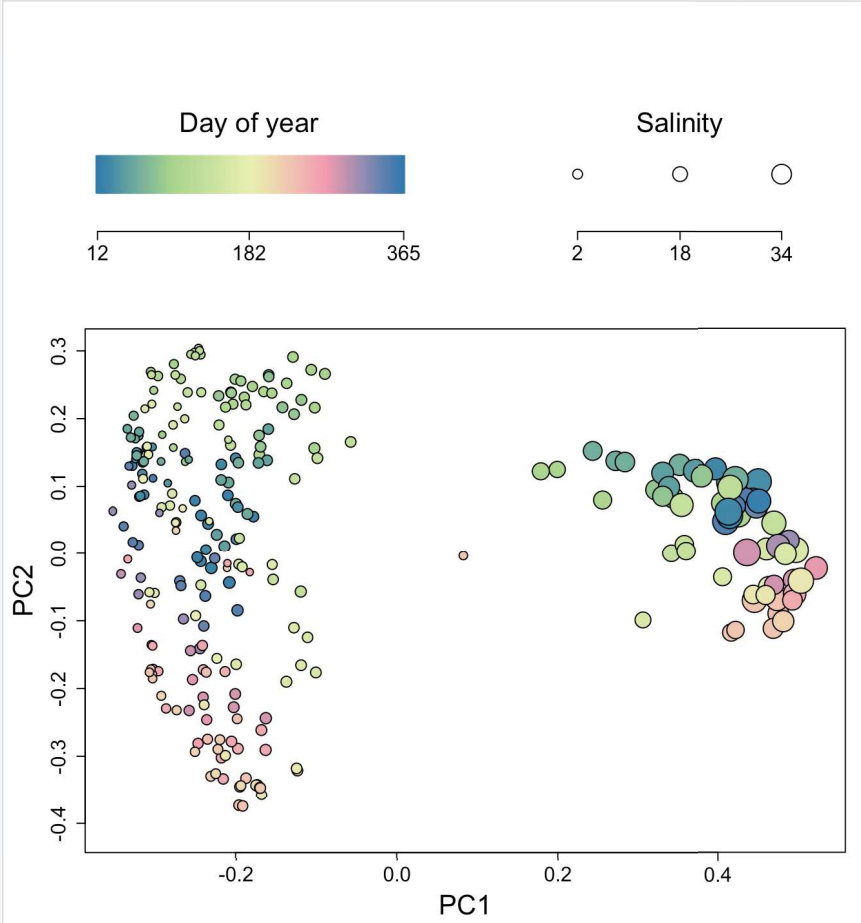
Import Dataset | 282 MiB | Grid

R | Global Environment

Name | Type | Length | Size | Value

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



RStudio

asvportal.R | merged_df$emo | merged$events | 16S_18S_combined.R | BS_MAG2_used_in_p

Source on Save | Run | Source

salin | Next | Prev | All | Replace | Replace | All

In selection | Match case | Whole word | Regex | Wrap

109:1 (Top Level) R Script

Console | Terminal | Background Jobs

R 4.3.2 · ~/aquatic/baltvib/analyses/

```
>
> # PCoA on Bray-Curtis distances
> pcoa <- pcoa(bray_dist, correction = "cailliez")
>
>
> # Plotting the color and size legends
> layout(matrix(c(1,2,3,3,3,3), 3, 2, byrow = T))
> par(mar=c(5,5,1,1), xpd = TRUE, cex.axis = 1.5)
> plot(1:365,rep(1,365), col = color_yday, pch = "|", cex = 3, axes = F, ylim = c(0.9,1.3), ylab = "",
xlab = "")
> axis(1, labels = c("12","182","365"), at =  c(1,182,365), cex = 3)
> text(182, 1.1, "Day of year", cex = 2)
> plot(c(0,10,20),rep(1,3), col = "black", pch = 1, cex = c(1.6, 2.4, 3.2), xlim = c(-5,25), ylim = c
(0.9,1.3), axes =F, ylab = "", xlab = "")
> axis(1, labels = c("2", "18", "34"), at =  c(0,10,20), cex = 3)
> text(10, 1.1, "Salinity", cex = 2)
>
>
> # Plotting the PCoA
> plot(pcoa$vectors[,1], pcoa$vectors[,2], pch = pch, col = "white", bg = "white", xlab = "PC1", ylab
= "PC2")
> # plotting the DS_2019 samples
> points(pcoa$vectors[DS_2019,1], pcoa$vectors[DS_2019,2], pch = pch[DS_2019], col = "black", bg = col
or_yday[yday[DS_2019]], cex = 1.0 + as.numeric(salinity[DS_2019])/10, xlab = xlab, ylab = ylab)
> # adding the DS_2013 samples (squares)
> points(pcoa$vectors[DS_2013,1], pcoa$vectors[DS_2013,2], pch = pch[DS_2013], col = "black", bg = col
or_yday[yday[DS_2013]], cex = 1.0 + as.numeric(salinity[DS_2013])/10, xlab = xlab, ylab = ylab)
>
>
>
>
>
>
>
>
>
>
>
>
>
```

Environment | History | Connections | Tutorial

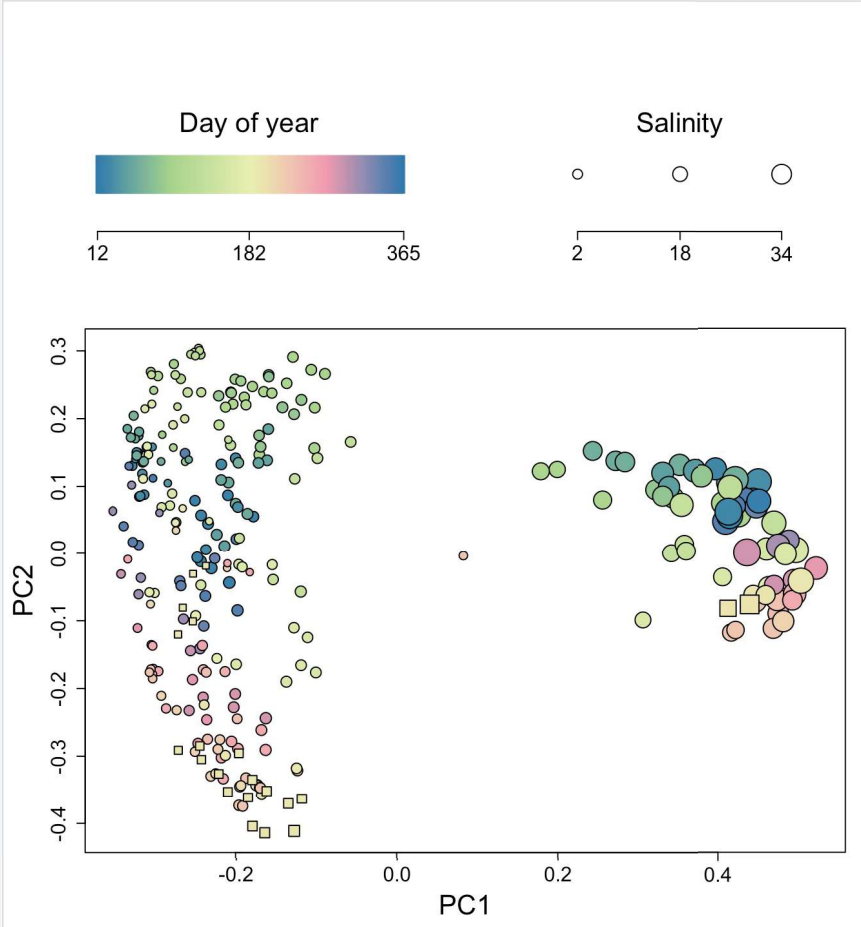
Import Dataset | 366 MiB | Grid

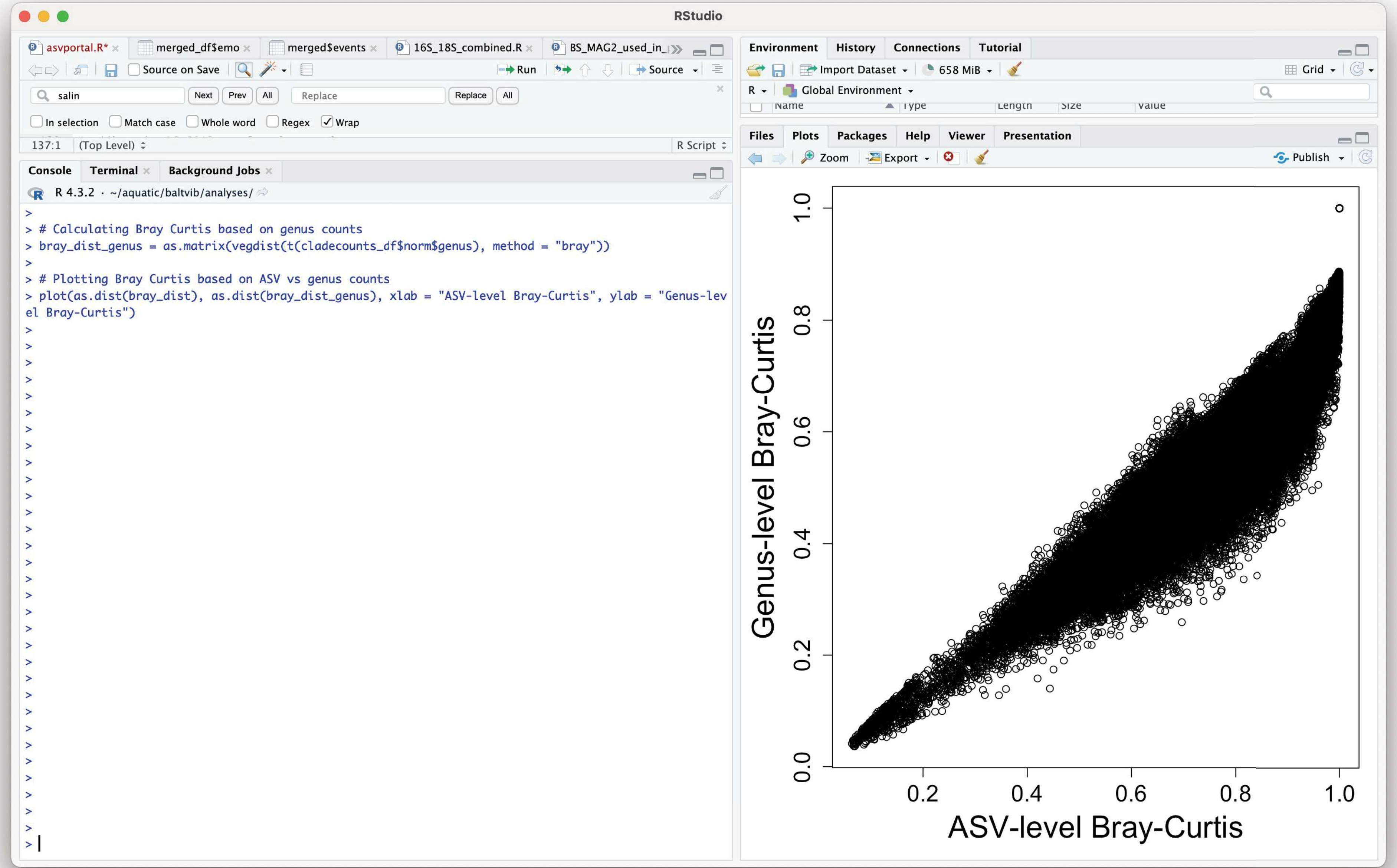
R | Global Environment

Name | Type | Length | Size | Value

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish




RStudio
asvportal.R*
merged_df$emo
merged$events
16S_18S_combined.R
BS_MAG2_used_in_
Source on Save
Run
Source
salin
Next
Prev
All
Replace
Replace
All
In selection
Match case
Whole word
Regex
Wrap
137:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/aquatic/baltvib/analyses/
> # Calculating Bray Curtis based on genus counts
> bray_dist_genus = as.matrix(vegdist(t(cladecounts_df$norm$genus), method = "bray"))
> # Plotting Bray Curtis based on ASV vs genus counts
> plot(as.dist(bray_dist), as.dist(bray_dist_genus), xlab = "ASV-level Bray-Curtis", ylab = "Genus-level Bray-Curtis")
Environment
History
Connections
Tutorial
Import Dataset
658 MiB
Grid
R
Global Environment
Name
Type
Length
Size
Value
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
0.0
0.2
0.4
0.6
0.8
1.0
Genus-level Bray-Curtis
0.2
0.4
0.6
0.8
1.0
ASV-level Bray-Curtis

Future plans for genetic data in SBDI

- Support more marker genes
- Integrate shotgun metagenomics data in the SBDI Bioatlas using the ENA MGnify pipeline
- (Long term) Integrate dataflow from NGI to SBDI

Thanks for listening!

More information: https://asv-portal.biodiversitydata.se

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

People involved: Maria Prager (KI/SU), Daniel Lundin (LnU), Jeanette Tångrot (Umu/NBIS), Tobias Anderman (UU), Anna Rosling (UU), Anders Andersson(KTH)