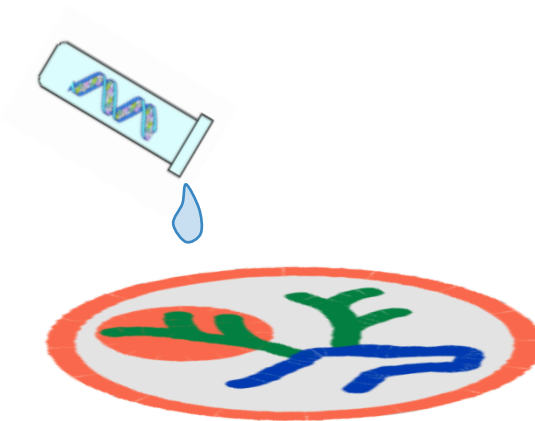




Integration of molecular data into Biodiversity Atlas Sweden (BAS)



Stockholm

Maria Prager (SU/KI)

Anders Andersson (KTH)

Kalmar

Diego Brambilla (LNU)

Daniel Lundin (LNU)

Uppsala

Jeanette Tångrot (UU/UmU)

Anna Rosling (UU)



Biodiversity Atlas Sweden (BAS)

Systematic monitoring data



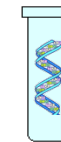
Biotelemetry (tracking) data



Natural history collection data



Molecular biodiversity data

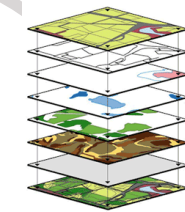


Metabarcoding
Metagenomics

python™

R
ALA4R...

OpenSci
Mirroreum: R online



Spatial
portal



Biodiversity Atlas Sweden (BAS)

Systematic monitoring data



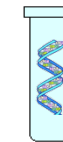
Biotelemetry (tracking) data



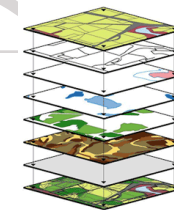
Natural history collection data



Molecular biodiversity data



Metabarcoding
Metagenomics



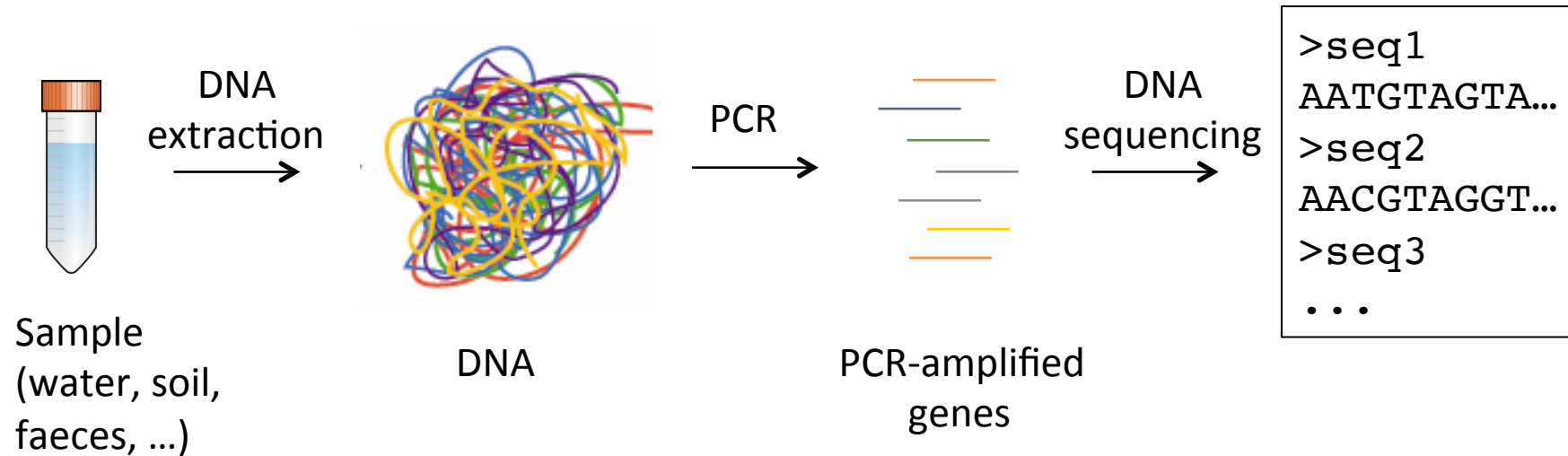
Spatial
portal



OpenSci
Mirroreum: R online

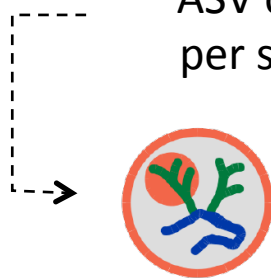


Metabarcoding (marker-gene amplicon sequencing)



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

ASV counts per sample



BAS

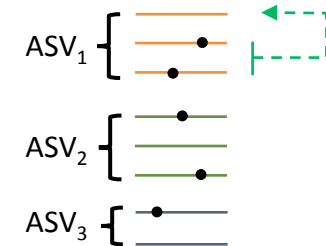
Matching of ASVs to reference database

←

with taxonomically classified sequences (e.g. Unite, Silva)

Error correction ↓ (denoising)

Amplicon Sequence Variants (ASVs)





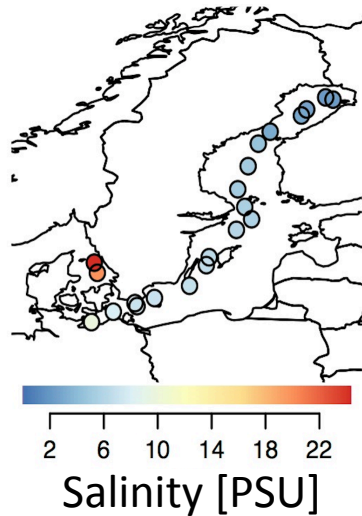
Organism groups / barcoding regions / databases

Organism group	Genomic region	Database
Metazoa	COI	BOLD
Fungi	ITS	Unite
Protists	18S	UniEuk/PR2
Prokaryotes	16S	SILVA or GTDB



Metabarcoding dataset example

Illumina sequencing of 21 surface water samples

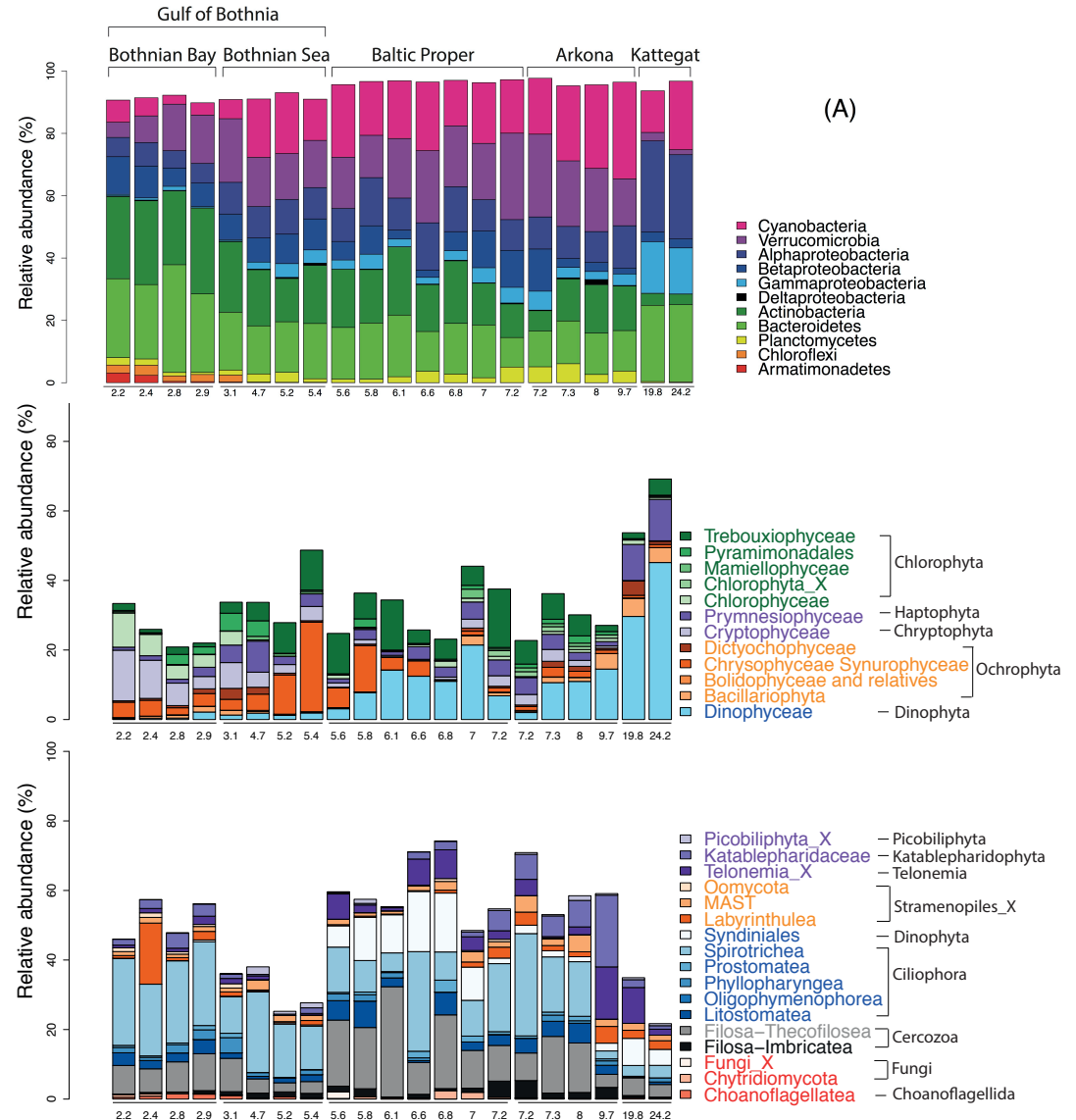


Bacteria (16S)

Eukaryotes (18S)

Phytoplankton

Heterotrophic protists





Biodiversity Atlas Sweden (BAS)

Sequencing provider



Lat:	>seq1
Lon:	AATGTAGTA...
Temp:	>seq2
Salinity:	AACGTAGGT...
...	...

Metadata

Sequences

Simplified
submission
of reads &
metadata

[Stockholm]



Sequence & metadata repositories





<https://enasubmitguide.readthedocs.io/en/latest/>

Apps Privat SU Scilife BAS GDocs GBIF ALA Docker BAS Git ENAgui » Other Bookmarks

🏠 The BAS guide to ENA submission

latest

Search docs

☰ Submitting metabarcoding data to ENA

- A guide from Biodiversity Atlas Sweden (BAS)

☰ Preparation for submission

Step 1: Prepare data and metadata

Step 2: Register with ENA

☰ 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: Create spreadsheet template

Step 3d: Edit spreadsheet structure

Step 3e: Add sample metadata

📖 Read the Docs

v: latest ▾

Docs » Submitting metabarcoding data to ENA

[🔗 Edit on GitHub](#)

This page is under construction

Submitting metabarcoding data to ENA

- A guide from Biodiversity Atlas Sweden (BAS)

This is a guide on how to submit sequence reads from environmental samples to the [European Nucleotide Archive \(ENA\)](#), provided by the [Biodiversity Atlas Sweden \(BAS\)](#) project. Our guide is largely a summary of [ENA's own extensive instructions](#), with added pointers on issues specific to submission of metabarcoding 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 only.

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 called *reads* and 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



Biodiversity Atlas Sweden (BAS)

Sequencing provider



Lat:	>seq1
Lon:	AATGTAGTA...
Temp:	>seq2
Salinity:	AACGTAGGT...
...	...

Metadata

Sequences

Simplified submission of reads & metadata

[Stockholm]

Sequence & metadata repositories



Pipeline for denoising & taxon. annotation of ASVs

[Kalmar, Uppsala]

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



Pipeline for
denoising & taxon.
annotation of ASVs

ILLUMINA (SHORT READ) DATA

[Kalmar]

Prokaryotes (16S)

Protists (18S)

Fungi (ITS)

Metazoa (COI)

PACBIO (LONG READ) DATA

[Uppsala]

Prokaryotes (16S)

Protists (18S)

Fungi (ITS)

Metazoa (COI)

nf-core/ampliseq

16S rRNA amplicon sequencing analysis workflow using QIIME2

build failing nextflow ≥18.10.1

install with bioconda docker build manual singularity available

nf-core collaboration.

Main developer Alexander Peltzer, Tuebingen

Pipeline under construction



Biodiversity Atlas Sweden (BAS)

Sequencing provider



Sequence & metadata repositories



Lat:	>seq1
Lon:	AATGTAGTA...
Temp:	>seq2
Salinity:	AACGTAGGT...
...	...

Metadata

Sequences

Simplified submission of reads & metadata

[Stockholm]

Pipeline/guide for submission of taxon. annotated ASV counts & metadata

[Stockholm]

Pipeline for denoising & taxon. annotation of ASVs

[Kalmar, Uppsala]

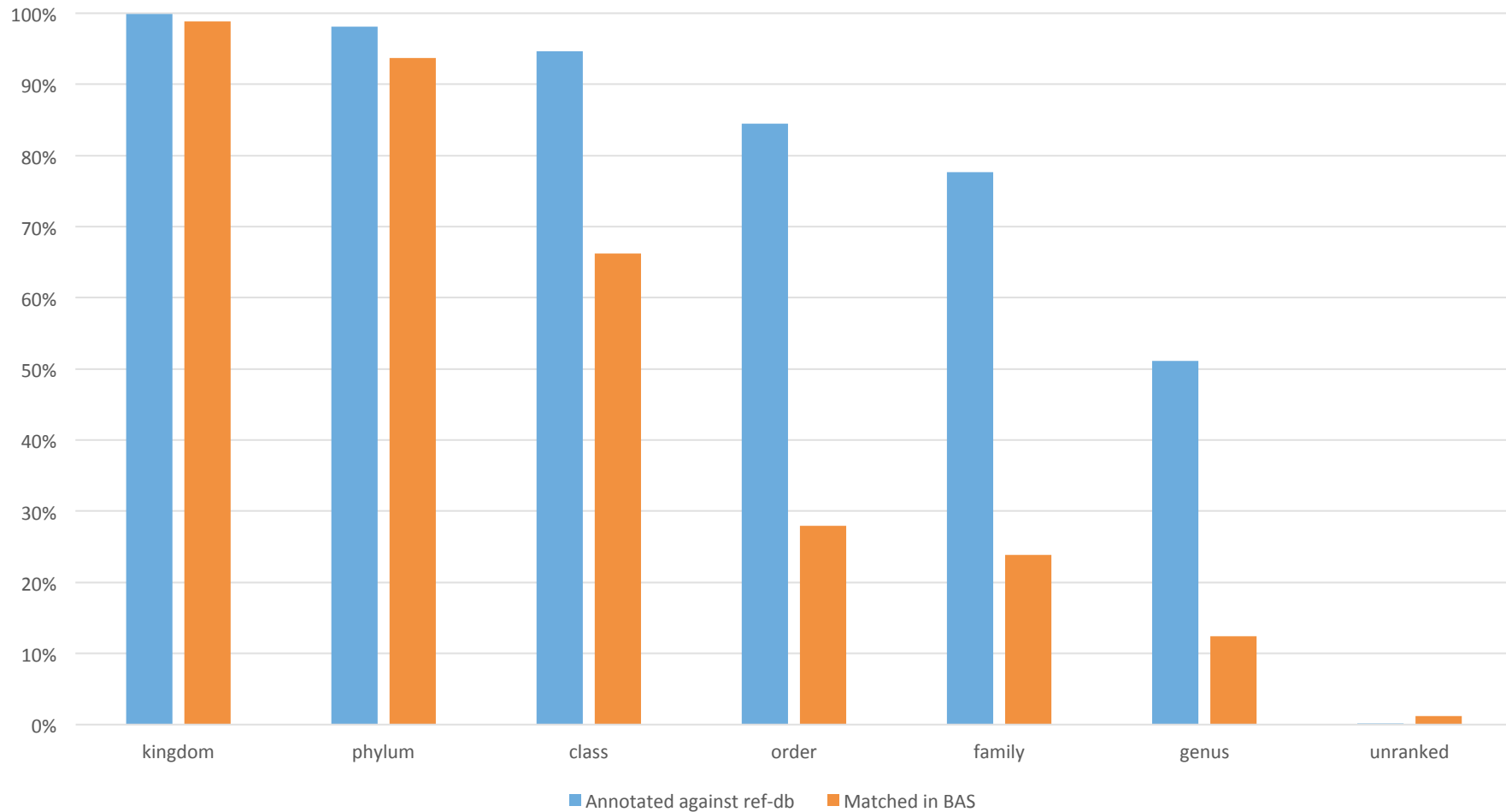


OTU	Species	Sample1	Sample2
1	E.coli	17	0
2	S.aurus	231	11800
3	unknown	30	0
...



Prokaryote taxonomy in SILVA db & BAS

Example: Percentage of SMHI Baltic Sea records (tot 2975) annotated against SILVA db vs. matching BAS taxonomy, at specified or lower taxonomic level



SILVA-GBIF collaboration -> first improvements expected 'over next few months'



Alternative data models for prokaryotes

1. No explicit ASV-IDs – ASVs uniquely identified via sequence + primer attributes in BAS

Occurrence core file

occurrenceID	scientificName	taxonRank	kingdom	phylum	class	order	family	genus
abc:256:11	Marine_Group_II	family	Archaea	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Marine_Group_II	
abc:256:12	Subgroup_17	class	Bacteria	Acidobacteria	Subgroup_17			

GGBN Amplification extension file (currently not available for search or analysis in BAS)

occurrenceID	primerSequenceForward	primerSequenceReverse	barcodeSequence
abc:256:11	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATC C	GCGCGAAAACCTTCACACTGCAGGAACTGTGATGAGGGAACTCA ...
abc:256:12	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATC C	TTAGGAATCTTCTGCAATGCGCGCAAGCGTGACAGATTGAGCCAG ...

2. Unique ASV-IDs (valid across samples and datasets), linked to sequences in external database

Occurrence core file (non-Linnean scientific names also not searchable in BAS)

occurrenceID	scientificName	taxonRank	kingdom	phylum	class	order	family	genus
abc:256:11	seq_1021	unranked	Archaea	Euryarchaeota	Thermoplasmata	Thermoplasmatales	Marine_Group_II	
abc:256:12	seq_2887	unranked	Bacteria	Acidobacteria	Subgroup_17			

Separate ASV database

asvID	primerSequenceForward	primerSequenceReverse	barcodeSequence
seq_1021	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATC C	GCGCGAAAACCTTCACACTGCAGGAACTGTGATGAGGGAACTCA ...
seq_2887	CCTACGGGNGGCWGCAG	GACTACHVGGGTATCTAATC C	TTAGGAATCTTCTGCAATGCGCGCAAGCGTGACAGATTGAGCCAG ...



Biodiversity Atlas Sweden (BAS)

Deliverable	2018	2019	2020	
D3.8 - A1	Submission service of sequencing- and metadata to INSDC databases	●	●	SU, UU, LnU, KTH, KI
D3.8 - A2	Service for mapping metabarcoding data to reference databases		●	SU, UU, LnU, KTH, KI
D3.8 - A3	Database of reference MOTUs for major Swedish biomes		●	SU, UU, LnU, KTH, KI
D3.8 - A4	Service for metagenome sequence annotation			SU, UU, LnU, KTH, KI
D3.8 - A5	Database with processed Swedish metagenome data			SU, UU, LnU, KTH, KI
D4.5 - A1	Integration of molecular data into ALA			● SU, UU, LnU, KTH, KI
D4.5 - A2	Analysis and visualization functionality for molecular data			SU, UU, LnU, KTH, KI
D4.5 - A3	Analysis module for community analysis based on molecular data			SU, UU, LnU, KTH, KI

Lat:	>seq1
Lon:	AATGTAGTA...
Temp:	>seq2
Salinity:	AACGTAGGT...
...	...

Metadata

Sequences

Simplified submission of reads & metadata

D3.8 - A1

[Stockholm]



Pipeline/guide for submission of taxon. annotated ASV counts & metadata

(D3.8 - A3)

D4.5 - A1

[Stockholm]



Pipeline for denoising & taxon. annotation of ASVs

D3.8 - A2

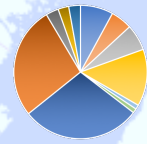
[Kalmar, Uppsala]

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



Future: Services for Analysis & Visualization

Visualisation of relative abundances



Aggregation at higher taxonomic levels

- Cyanobacteria
- Verrucomicrobia
- Alphaproteobacteria
- Betaproteobacteria
- Gammaproteobacteria
- Deltaproteobacteria
- Actinobacteria
- Bacteroidetes
- Planctomycetes
- Chloroflexi

