

Apscale manual

Last updated: 11.07.2025, apscale version 4.0.5

Author: Dominik Buchner

Table of contents

1. Introduction.....	3
2. Installation	3
3. Usage	5
3.1. Creating a project.....	5
3.2. Configuring the settings.....	5
3.3. Adding data	5
3.4. Running a module – Linear workflow	5
3.4.1. Paired-end merging	5
3.4.2. Primer trimming	5
3.4.3. Quality filtering	5
3.4.4. Dereplication.....	5
3.5. Running a module – Modular workflow	5
3.5.1. Denoising.....	5
3.5.2. Swarm clustering	5
3.5.3. Replicate merging.....	5
3.5.4. Negative control filtering	5
3.5.5. Read table generation	5
3.6. Working with the read data store.....	5
3.6.1. Apscale analyze.....	5
3.6.2. Adding sample metadata	5
3.6.3. Adding sequence metadata	5
3.6.4. Correcting species names via GBIF	5
3.6.5. Validating species occurrence via GBIF.....	5
3.6.6. Uploading datasets to ENA.....	5
3.6.7. Exporting read tables.....	5

1. Introduction

Apscale is a metabarcoding pipeline that handles the most common tasks in metabarcoding pipelines like paired-end merging, primer trimming, quality filtering, denoising, swarm and threshold-based clustering as well as basic data handling operations such as replicate merging and the removal of reads found in the negative controls. It uses a simple command line interface and is configured via a single configuration file. To add metadata to the dataset, a simple, browser-based interface has been introduced in version 4.0. Apscale automatically uses the available resources on the machine it runs on while still providing the option to use less if desired. All modules can be run on their own or as a comprehensive workflow.

Several different programs are called within the workflow. These include vsearch (Rognes et al. 2016), cutadapt (Martin 2011) and swarm (Mahé et al. 2021). Please cite those accordingly, when using apscale. A DuckDB backend was introduced in version 4.1. Please also cite the authors of DuckDB, they build an amazing tool that makes the current developments in Apscale possible.

Apscale (Buchner et al. 2022) has also been published and we are happy if we are cited too.

2. Installation

Apscale can be installed on all common operating systems (Windows, Linux, MacOS). Apscale requires Python 3.11 or higher and can be easily installed via pip in any command line:

```
pip install apscale
```

To update apscale run:

```
pip install --upgrade apscale
```

Apscale calls vsearch as well as swarm for multiple modules. Both programs should be installed and be in PATH to be executed from anywhere on the system.

Check the vsearch and swarm Github pages for further info:

<https://github.com/torognes/vsearch>

<https://github.com/torognes/swarm>

To check if every is correctly set up, please type this into your command line:

```
vsearch --version
```

```
swarm --version
```

It should return messages similar to these:

```
vsearch v2.30.0_win_x86_64, 127.9GB RAM, 24 cores  
https://github.com/torognes/vsearch
```

```
Rognes T, Flouris T, Nichols B, Quince C, Mahe F (2016)  
VSEARCH: a versatile open source tool for metagenomics  
PeerJ 4:e2584 doi: 10.7717/peerj.2584 https://doi.org/10.7717/peerj.2584
```

```
Compiled with support for gzip-compressed files, and the library is loaded.  
zlib version 1.2.13, compile flags 65  
Compiled with support for bzip2-compressed files, and the library is loaded.
```

```
Swarm 3.1.5  
Copyright (C) 2012-2024 Torbjorn Rognes and Frederic Mahe  
https://github.com/torognes/swarm
```

```
Mahe F, Rognes T, Quince C, de Vargas C, Dunthorn M (2014)  
Swarm: robust and fast clustering method for amplicon-based studies  
PeerJ 2:e593 https://doi.org/10.7717/peerj.593
```

```
Mahe F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015)  
Swarm v2: highly-scalable and high-resolution amplicon clustering  
PeerJ 3:e1420 https://doi.org/10.7717/peerj.1420
```

```
Mahe F, Czech L, Stamatakis A, Quince C, de Vargas C, Dunthorn M, Rognes T (2022)  
Swarm v3: towards tera-scale amplicon clustering  
Bioinformatics 38:1, 267-269 https://doi.org/10.1093/bioinformatics/btab493
```

Further dependencies - cutadapt

Apscale also calls cutadapt with the primer trimming module. Cutadapt should be downloaded and installed automatically with the Apscale installation. To check this, type:

```
cutadapt --version
```

and it should return the version number, for example:

```
5.1
```

3. Usage

- 3.1. Creating a project
- 3.2. Configuring the settings
- 3.3. Adding data
- 3.4. Running a module – Linear workflow
 - 3.4.1. Running the full pipeline
 - 3.4.2. Paired-end merging
 - 3.4.3. Primer trimming
 - 3.4.4. Quality filtering
 - 3.4.5. Dereplication
- 3.5. Running a module – Modular workflow
 - 3.5.1. Denoising
 - 3.5.2. Swarm clustering
 - 3.5.3. Replicate merging
 - 3.5.4. Negative control filtering
 - 3.5.5. Read table generation
- 3.6. Working with the read data store
 - 3.6.1. Apscale analyze
 - 3.6.2. Adding sample metadata
 - 3.6.3. Adding sequence metadata
 - 3.6.4. Correcting species names via GBIF
 - 3.6.5. Validating species occurrence via GBIF
 - 3.6.6. Uploading datasets to ENA
 - 3.6.7. Exporting read tables