

skovutils

Small utility helpers for phylogenetic clustering, repeated-measures correlation, and abundance filtering used in microbiome analysis workflows.

Installation

Install straight from GitHub (PEP 517 build via [pip](#)).

```
pip install git+https://github.com/The-Skov/skovutils.git
```

The package depends on [numpy](#), [numba](#), [pandas](#), [scipy](#), and [scikit-bio](#). Optional examples may also use [pingouin](#).

Usage

Import the package and check the version:

```
import skovutils
print(skovutils.__version__)
```

Example: filtering with [minmin](#)

This example demonstrates the [minmin](#) function from [skovutils.filtering](#). It keeps observations (rows) that have at least [abundance](#) in a minimum fraction of samples given by [prevalence](#).

```
from skovutils.filtering import minmin
from skbio import Table
import numpy as np

data = np.array([
    [0.0, 0.02, 0.03], # obs1: low abundance in 2/3 samples
    [0.1, 0.0, 0.0],   # obs2: abundant in 1/3
    [0.05, 0.06, 0.07], # obs3: abundant in all samples
    [0.0, 0.0, 0.0],   # obs4: absent everywhere
])
obs_ids = ['obs1', 'obs2', 'obs3', 'obs4']
samp_ids = ['samp1', 'samp2', 'samp3']
table = Table(data, obs_ids, samp_ids)

# Keep observations that have abundance >= 0.05 in at least 50% of samples
filtered = minmin(table, abundance=0.05, prevalence=0.5)

print("Original observations:", list(table.ids(axis='observation')))
print("Filtered observations:", list(filtered.ids(axis='observation')))
```

```
# Expected output:  
# Original observations: ['obs1', 'obs2', 'obs3', 'obs4']  
# Filtered observations: ['obs3']
```

License

Released under the MIT License. See [LICENSE](#) for details.