

## amazon-continuation

A new look at the Amazon Continuum Metagenomes

Author: Kai Blumberg

Email: [kblumberg@email.arizona.edu](mailto:kblumberg@email.arizona.edu)

Script: `amazon_xml_parser.py`

A script to parse the Amazon Continuum Metagenomes project's metadata from an xml file. The script parses all the samples into a list of dictionaries, each of which contains all metadata for a sample as key value pairs in the dictionaries.

After parsing the metadata, the script plots depth profiles of biogeochemical parameters such as dissolved oxygen and dissolved inorganic carbon.

The xml file was obtained from the NCBI National center for Biotechnological information bioproject 237344 and clicking on **send to** chose **file** then chose select the **Format** as **Full XML (text)** then click **Create File**.

XML parsing script based on the example provided by Ken Youens-Clark from the University of Arizona Biosystems analytics course, from lecture 9 as well as the `xml.etree.ElementTree` documentation.

In order to run the script the user will need to have matplotlib installed. If using an anaconda environment simply run `conda install -c conda-forge matplotlib`.

Run the program by calling:

```
./amazon_xml_parser.py biosample_result.xml
```

The program should generate the output files: `Dissolved_Oxygen_profile.png` and `Dissolved_Inorganic_carbon_profile.png`

Test that the program has executed correctly by running: `make test`