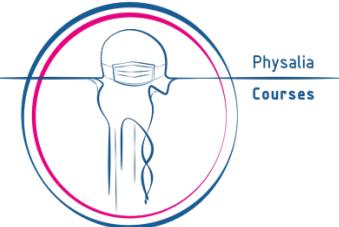
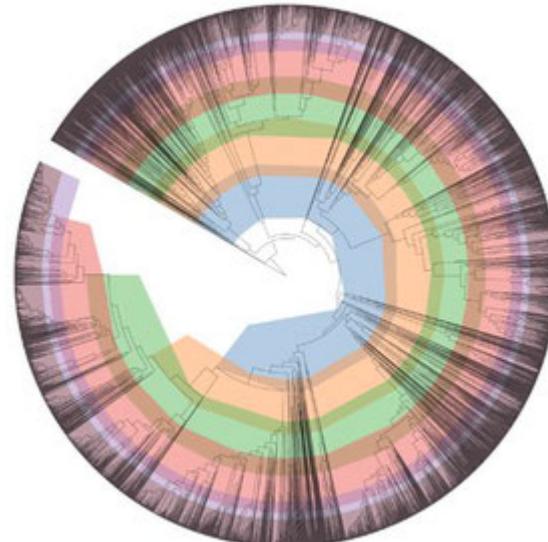


ENVIRONMENTAL METAGENOMICS

Physalia course, online, 11-15 November 2024

Course outline and practical information

Nikolay Oskolkov, Lund University, NBIS SciLifeLab
Luis Pedro Coelho, Queensland University of Technology



NB: original course material courtesy:

Dr. Antti Karkman, University of Helsinki

Dr. Igor Pessi, Finnish Environment Institute (SYKE)

About us

Organizer: Carlo Pecoraro, Physalia courses

info@physalia-courses.org



Instructors:

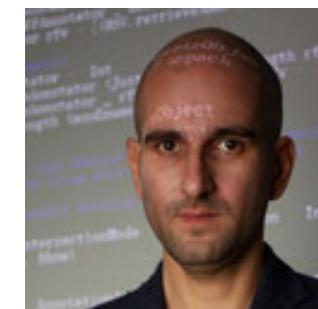
Dr. Nikolay Oskolkov, Lund University, NBIS SciLifeLab

Nikolay.Oskolkov@biol.lu.se



Dr. Luis Pedro Coelho, Queensland University of Technology

luis@luispedro.org





@NikolayOskolkov



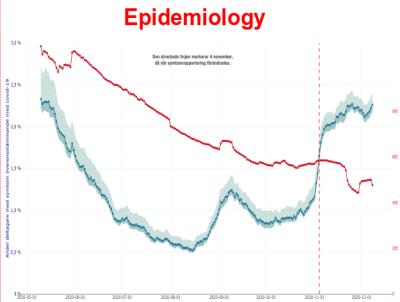
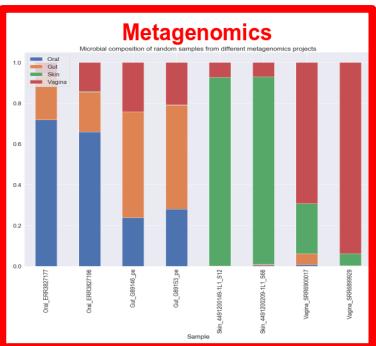
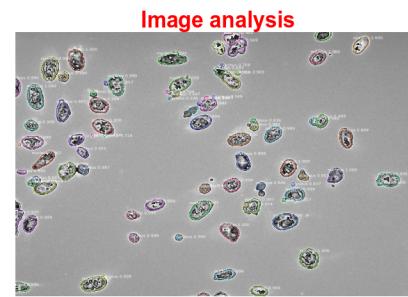
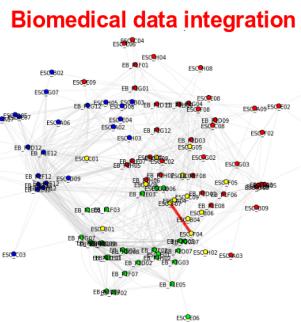
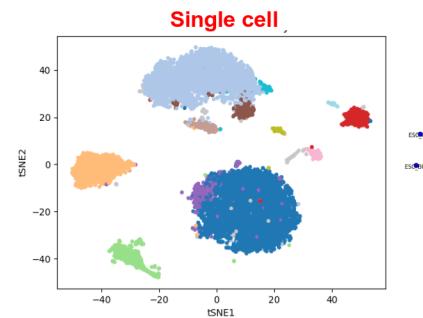
Personal homepage:
<https://nikolay-oskolkov.com>

Brief introduction: who am I

2007 PhD in theoretical physics

2011 medical genetics at Lund University

2016 working at NBIS SciLifeLab, Sweden



Pochon et al. *Genome Biology* (2023) 24:242
<https://doi.org/10.1186/s13059-023-03083-9>

METHOD

Genome Biology

Open Access



aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow

Zoé Pochon^{1,2†}, Nora Bergfeldt^{1,3,4†}, Emrah Kirdök⁵, Mário Vicente^{1,2}, Thijessen Naidoo^{1,2,6,7}, Tom van der Valk^{1,4}, N. Ezgi Altıntışık⁸, Małgorzata Krzewińska^{1,2}, Love Dalén^{1,3}, Anders Götherström^{1,2†}, Claudio Mirabello^{9†}, Per Unneberg^{10†} and Nikolay Oskolkov^{11†}

¹Zoé Pochon, Nora Bergfeldt, Anders Götherström, Claudio Mirabello, Per Unneberg, and Nikolay Oskolkov shared authorship.

*Correspondence:
Nikolay.Oskolkov@biol.lu.se

¹¹Department of Biology,
Science for Life Laboratory,
National Bioinformatics
Infrastructure Sweden, Lund
University, Lund, Sweden
Full list of author information is
available at the end of the article

Abstract

Analysis of microbial data from archaeological samples is a growing field with great potential for understanding ancient environments, lifestyles, and diseases. However, high error rates have been a challenge in ancient metagenomics, and the availability of computational frameworks that meet the demands of the field is limited. Here, we propose aMeta, an accurate metagenomic profiling workflow for ancient DNA designed to minimize the amount of false discoveries and computer memory requirements. Using simulated data, we benchmark aMeta against a current state-of-the-art workflow and demonstrate its superiority in microbial detection and authentication, as well as substantially lower usage of computer memory.

Keywords: Ancient metagenomics, Pathogen detection, Microbiome profiling, Ancient DNA

Background

Historically, ancient DNA (aDNA) studies have focused on human and faunal evolution and demography, extracting and analyzing predominantly eukaryotic aDNA [1–3]. With the development of next-generation sequencing (NGS) technologies, it was demonstrated that host-associated microbial aDNA from eukaryotic remains, which was previously treated as a sequencing by-product, can provide valuable information about ancient pandemics, lifestyle, and population migrations in the past [4–6]. Modern technologies have made it possible to study not only ancient microbiomes populating eukaryotic hosts, but also sedimentary ancient DNA (sedaDNA), which has rapidly become an independent branch of palaeogenetics, delivering unprecedented information about hominin and animal evolution without the need to analyze historical bones and teeth [7–12]. Previously available in microbial ecology, meta-barcoding methods lack validation and authentication power, and therefore, shotgun metagenomics has become the *de facto* standard in ancient microbiome research [13]. However, accurate detection,

README MIT license

aMeta: an accurate and memory-efficient ancient Metagenomic profiling workflow

snakemake ≥6.16.0 Tests passing

About

aMeta is a Snakemake workflow for identifying microbial sequences in ancient DNA shotgun metagenomics samples. The workflow performs:

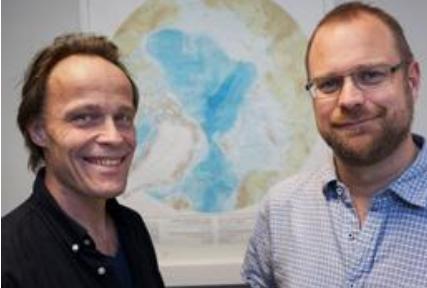
- trimming adapter sequences and removing reads shorter than 30 bp with Cutadapt
- quality control before and after trimming with FastQC and MultiQC
- taxonomic sequence kmer-based classification with KrakenUniq
- sequence alignment with Bowtie2 and screening for common microbial pathogens
- deamination pattern analysis with MapDamage2
- Lowest Common Ancestor (LCA) sequence alignment with Malt
- authentication and validation of identified microbial species with MaltExtract

When using aMeta and / or pre-built databases provided together with the workflow for your research projects, please cite our preprint: <https://www.biorxiv.org/content/10.1101/2022.10.03.510579v1>

Authors

- Nikolay Oskolkov (@LeandroRitter) nikolay.oskolkov@scilifelab.se
- Claudio Mirabello (@clami66) claudio.mirabello@scilifelab.se
- Per Unneberg (@percyfal) per.unneberg@scilifelab.se

<https://github.com/NBISweden/aMeta>



About you

- Name
- University/Institute/Company
- Research interest(s)
- Previous experience(s) with microbial ecology, metagenomics, bioinformatics, etc.
- General hopes for this course



Course outline

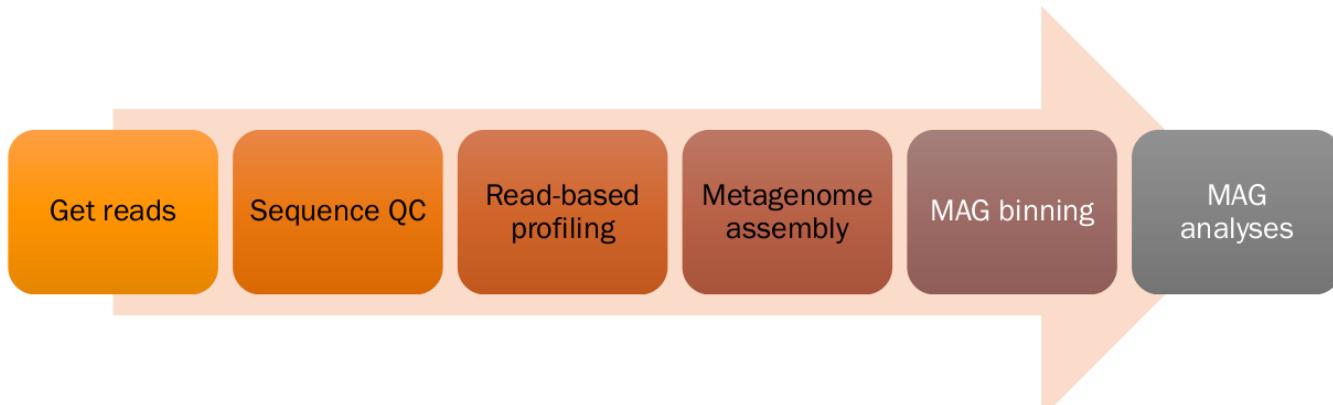
Day 1: introduction, setting up, connecting to server, getting raw data, exploring data

Day2: quality control, adapter removal, read-based taxonomic classification

Day3: *de-novo* assembly, taxonomic profiling and abundance quantification of contigs

Day4: advanced metagenome assembly, quality control of assembled metagenome

Day5: metagenomic binning and functional analysis of MAGs



Practical information: GitHub and Zoom

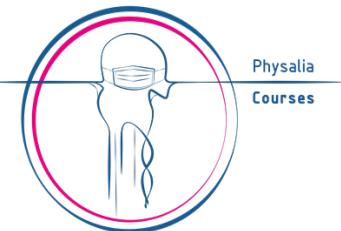
The course will take place **in Zoom from 9am to 1pm (CET)**

Links to the Zoom room will be posted in Slack

The course GitHub repository containing lectures and exercises is:

https://github.com/NikolayOskolkov/Physalia_EnvMetagenomics_2024

Please bookmark this address!



Practical information: Amazon Cloud (AWS EC2)



We will use the Cloud Computing service from Amazon, which we will access via **ssh** (secure shell protocol)

https://github.com/NikolayOskolkov/Physalia_EnvMetagenomics_2024/blob/main/exercises.md#setting-up-the-cloud-computing

See [here](#) for information on how to connect, but remember:

- The IP address changes every day
- Everyone is given a username, with a **home** and **shared** folders
 - List of usernames can be found in Slack
 - The **shared** folder is copy-only: do not delete, move, rename, or write



Practical information: conda



System for software management (python, R, JavaScript, C++, ...)

Allows easy installation of software in dedicated environments,
separated from the main environment and other conda environments

- The environments that we will use have been already set up for everyone

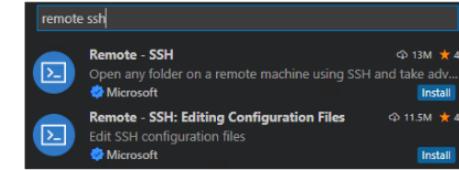
General conda commands

```
> conda env list  
> conda activate ambiente  
> conda deactivate
```



Practical information: setting up VS Code

- Download and install VS Code: code.visualstudio.com/Download
- Launch VS Code
- Go to [View -> Extensions](#)
- Search for and install the extension [Remote-SSH](#)
- See [here](#) for a step-by-step guide on how to connect to the Amazon Cloud



https://github.com/NikolayOskolkov/Physalia_EnvMetagenomics_2024/blob/main/exercises.md

Samples to work on during this course

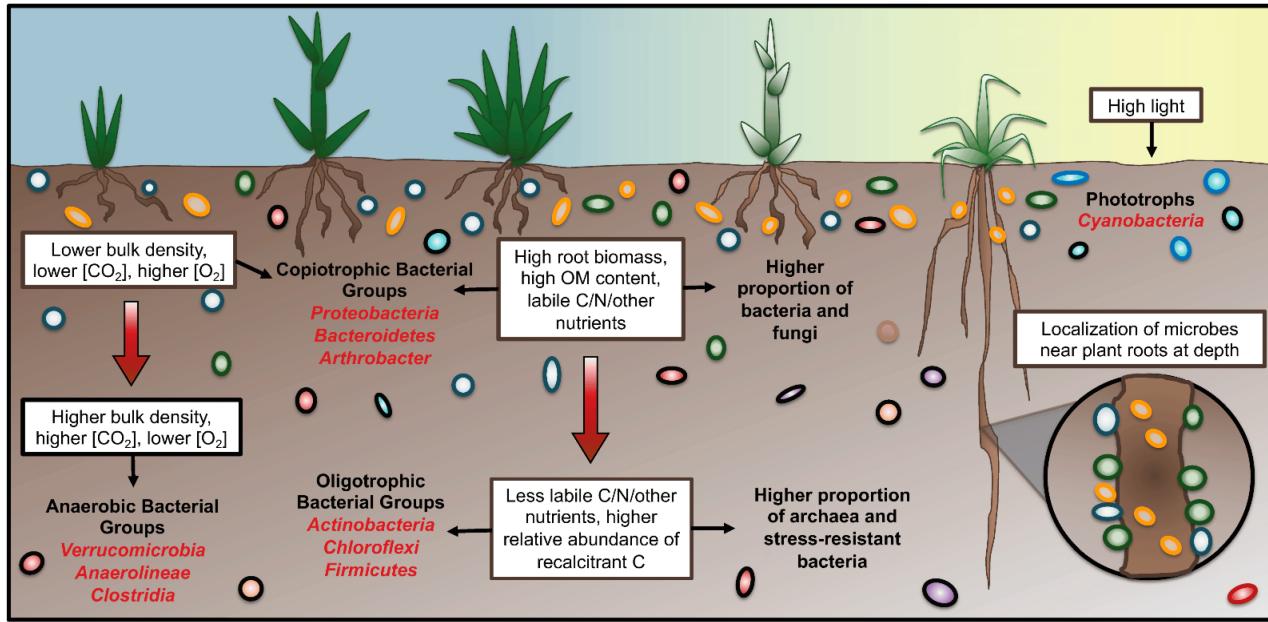
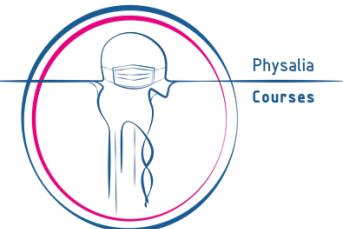


Image adapted from Naylor et al., Microorganisms 2022

The 5 samples are collected from **urban soil (Fudan University, Shanghai)**. Sample 3 and sample 4 are duplicate samples. Urban soil is formed by the occupation of natural soil by cities and the strong influence of human activities. As soil is an important reservoir of antibiotic resistance genes in the environment. Exploring soil resistance genes is becoming increasingly important for soil safety and human health.



Illumina (ILM) + Nanopore (ONT) samples are available at ~/Share