Syllabus of the 2015 Summer Course on Theoretical and Practical Approaches to Metagenomics and Viral discovery

Course Description:

This course introduces fundamental concepts of metagenomics experimental design and data analysis using next-generation sequencing data. Topics covered include rational for experimental design of metagenomics experiments, theoretical basis of the most common bioinformatics and statistical tools used for metagenome analysis, and practical approaches for data analysis using web-based and command line tools and commands. Theoretical and practical classes will be provided. The course will also exemplify some of the advances in microbial ecology and viral discovery derived from recent metagenomic studies, including some mainstream approaches and a practical session using the GenSeed-HMM program.

Instructors:

Alejandro Reyes, E-mail: a.reyes@uniandes.edu.co
Department of Biological Sciences, Universidad de los Andes, Bogotá, Colombia / Center for Genome Sciences and Systems Biology, Department of Pathology and Immunology, Washington University in Saint Louis, USA

Arthur Gruber, E-mail: argruber@usp.br
Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, Brazil

Schedule and location:

February 1 to 4, 2016
9 am to 12 pm and 2 pm to 5 pm
Institute of Biomedical Sciences, Edifício Biomédicas IV, rooms 3 and 5, University of São Paulo
Class activities: 24 h
Home activities: 6 h

Prerequisites:

Practical sessions will be taught using Linux servers. Knowledge of the Linux command Line is required for the practical classes. Graduate and undergraduate students are welcome.

Application:

Please send your CV and a short letter of intent to Prof. Arthur Gruber (argruber@usp.br). Total number of seats: 20.

Language:

The course will be taught in English.
Bibliography:

No textbook is required for this course. Some papers covering the main topics are listed below. Additional papers will be assigned and made available on the course’s web site in advance.


Table of content

Day 1

Morning – Theoretical Class

Introduction to Bioinformatics analysis (1.5h)

- Biological questions x Bioinformatics analyses
- Main bioinformatics approaches
- Pairwise alignments
- Sequence databases
- Orthology and pathway databases

DNA sequencing and next-generation sequencing (1.5 h)

- Sanger sequencing
- Next-generation sequencers (454 and Illumina)
- Pros and cons of each platform
- Data formats (FASTA, FASTQ, SFF) and interconversion tools
- Data processing: data quality, quality filtering, end trimming
- DNA assembly: programs, evaluation of assembly (N50, accumulative contig size, etc.)

Afternoon – Practical Class

DNA sequencing and next-generation sequencing (3.0 h)

- Quality evaluation and adapter identification – FastQC
- Quality filtering, end trimming and adaptor removal – cutadapt
- Data assembly of short reads (Illumina) – ABySS
- Data assembly of long reads (454) - Newbler
- Read mapping – Bowtie
- Data visualization – Tablet or IGV

Day 2

Morning – Theoretical Class

Introduction to Metagenomics (1.5 h)

- What is metagenomics
- Types of metagenomes: Amplicon, Shotgun, Functional
- Amplicon metagenomics: History, phylogenetic marker, examples
- Shotgun metagenomics: History and examples
- Functional metagenomics: Examples
Amplicon metagenomics (1.5 h)

- Experimental design
- Preliminary data analysis
  - Quality Control
  - Denoising, Filtering, Trimming
  - Chimeras
- Clustering: OTU calling
- Taxonomical assignment
- The OTU table
- Multivariate statistical analysis: Alpha and Beta diversity, PCoA
- The qiime software package

Afternoon – Practical Class

Amplicon metagenomics (3.0 h)

- Data analysis and applications
- Using QIIME and other amplicon analysis software
- Galaxy graphical interface.
- Using QIIME in Galaxy
- Other tools

Day 3

Morning – Theoretical Class

Shotgun metagenomics (3.0 h)

- Experimental design
- Advantage and disadvantages
- Relative vs absolute analysis
- Functional and taxonomical annotation of metagenomes
- Metagenome assembly
- Viral metagenomics

Afternoon – Practical Class

Shotgun metagenomics (3.0 h)

- Analysis and annotation of shotgun metagenomes
- Functional analysis
- Taxonomical analysis
- Metagenome assembly
Day 4

Morning – Theoretical Class

Classical viral discovery (1.0h)

- Classical methods and paradigms on viral discovery
- Running a conventional assembly on metagenomic data

Innovative methods for viral discovery (2.5h)

- Innovative approaches for viral discovery from metagenomic data
- VirSorter for viral signal detection
- Alternative sequence reconstruction methods
- Viral databases of orthologous groups and profile HMMs: vFam, POGs and NCVOGs (sections of 2014 update of COGs), Viral OGs (viral section of eggNOG)
- Reconstructing viral genomes using profile HMMs as seeds

Afternoon – Practical Class

Viral discovery and annotation of *Alpavirinae* using GenSeed-HMM (3.0 h)

- Data pre-processing
- Running a conventional assembly on metagenomic data
- Reconstructing viral genomes using profile HMMs as seeds
- Post-assembly comparative analyses
- Automated sequence annotation with EGene 2 platform