

Introduction to RADseq: bioinformatics

About the course

Time: October 7 – 11, 2013

Place: The Natural History Museum and the Centre for Ecological and Evolutionary Synthesis, University of Oslo.

Course credits: 2.5 ECTS

Assessment: Written report within two weeks after course is finished.

Registration: Register before September 11, 2013 at <https://nettskjema.uio.no/answer/56020.html>

Course fee: No course fee.

Maximum number of participants: For practical reasons the number of participants is limited to 30. If more students apply we will prioritize (1) participants from the course Introduction to RADseq: part I, (2) ForBio members and associates having generated RADseq data and need to analyze it, (3) ForBio members and associates with concrete plans to apply RADseq techniques in their research, (4) other ForBio members, and (5) other associates. Please describe how RADseq fits in your research plans and the stage you currently are at when you register for the course.

Teachers: Julian Catchen (jcatchen@uoregon.edu), Emiliano Trucchi (emiliano.trucchi@ibv.uio.no)

Anna Mazzarella (a.v.b.mazzarella@ibv.uio.no) Magnus Popp (magnus.popp@nhm.uio.no)

Prerequisites: You will need a laptop throughout the course. Make sure you have administrator privileges and can install software on the computer. Participants need a working knowledge of model based phylogenetic and/or population genetic analyses. No programming experience is required but participants need to be familiar with the UNIX environment and executing commands in a Bash shell.

A one-day introduction to UNIX, Bash, and Python is arranged Sunday October 3 if necessary. It may also be possible to attend the ForBio course “Introduction to bioinformatics for biosystematics” (<http://www.forbio.uio.no/events/courses/2013/bioinfointro2013.html>), please contact Magnus Popp (magnus.popp@nhm.uio.no) if you want to register.

Travel and accommodation: The course is arranged by the Research School in Biosystematics – ForBio – and travel and accommodation is refunded for ForBio members registered at other universities than University of Oslo. ForBio may also support master students registered at Danish, Swedish, or Norwegian universities who fulfill the prerequisites and need the course for their thesis work. ForBio will not refund travel or accommodation costs for other associates. See www.forbio.uio.no for more information on ForBio and membership.

Course content

Restriction site Associated DNA (RAD) sequencing, along with advances in sequencing technology, is making it feasible to obtain genome scale data not only from model organisms, but also from species without available genomic resources. RADseq techniques have been used for SNP discovery and genotyping, genotype-phenotype association mapping, linkage mapping, QTL analysis, hybridization and gene flow analysis, and population genetics. More recently, RADseq data has also been used to address problems in phylogeographic and phylogenetic studies by utilizing information from linked SNPs in “mini-contigs” obtained from paired end sequencing. RADseq is a flexible method for people looking to work on both model and non-model organisms.

The course will introduce the participants to Stacks (<http://creskolab.uoregon.edu/stacks/>), a bioinformatic pipeline for analyzing different types of RADseq data. Topics covered during the course include processing the raw data, assembling the sequences *de novo* and with a reference genome, SNP calling, and calculating basic population genetic statistics from the assembled data. The participants will also learn to explore the data, to look for signatures of natural selection, and to extract data for subsequent analyses. Some phylogenetic and phylogeographic analyses will be discussed. Some of the exercises will use data generated during the course Introduction to RADseq: part I, with which we will study the phylogeny of and putative hybridization among a group of cichlid fishes.

Tentative course schedule:

Monday October 7:

- 9:00-10:00: Welcome and general Introduction to the course (Julian and Emiliano)
- 10:00-11:00: Lecture: Introduction to RADseq (Emiliano)
- 11:00-12:00: Setting up our computers (Emiliano and Julian)
- 12:00-13:00: Lunch
- 13:00-17:00: Tutorial: Cleaning and demultiplexing RADseq data using *Stacks* (Julian and Emiliano)
- 17:00-18:00: Optional extra help with the day's tutorials

Tuesday October 8:

- 9:00-11:00: Lecture: RADseq data in nonmodel organisms, making a genetic map (Emiliano)
- 11:00-12:00: Student presentations (5)
- 12:00-13:00: Lunch
- 13:00-17:00: Tutorial: RADseq data in nonmodel organisms, making a genetic map using *Stacks* (Julian and Emiliano)
- 17:00-18:00: Optional extra help with the day's tutorials

Wednesday October 9:

- 9:00-11:00: Tutorial: Aligning RADseq data to a reference genome using *Stacks* (Julian and Emiliano)
- 11:00-12:00: Lecture: Examples of analyses you can do with a reference genome (Julian/Anna)
- 12:00-13:00: Lunch
- 13:00-14:00: Student presentations (5)
- 14:00-17:00: Tutorial: Population analyses using *Stacks* (Julian and Emiliano)
- 17:00-18:00: Optional extra help with the day's tutorials

Thursday October 10:

- 9:00-11:00: Tutorial: Exporting data from *Stacks* (Julian and Emiliano)
- 11:00-12:00: Lecture: Phylogenomics, phylogenetics and phylogeography (Magnus)
- 12:00-13:00: Lunch
- 13:00-14:00: Lecture: Using data from Stacks for phylogeography (Emiliano)
- 14:00-17:00: Tutorial: Phylogenetics and phylogeography using data from *Stacks* (Emiliano)

17:00-18:00: Optional extra help with the day's tutorials

Friday October 11:

9:00-10:00: Lecture: RADseq experimental setup considerations (Emiliano)

10:00-11:00: Student presentations (5)

11:00-12:00: Lecture: Knowing your data: NGS data from a technological perspective (NN)

12:00-13:00: Lunch

13:00-14:00: Student presentations (5)

14:00-15:00: Friday Seminar: The development of *Stacks* (Julian)

15:00-17:00: Open discussion/flexible time for help with past tutorials

17:00-18:00: Optional extra help with the day's tutorials