Phylogenomics

About the workshop

Time: Week I Nov 14 – 18, 2016, Week II March/April 2017.

Place: The Auditorium of the Lids hus, Natural History Museum, University of Oslo

Course credits: 5 ECTS

Assessment: Written report on provided case studies within two weeks after each week of the workshop is finished.

Registration: Registration before August 1st 2016

Fee: No fee for ForBio members or associates.

Maximum number of participants: 26

Teachers: Torsten Hugo Struck (t.h.struck@nhm.uio.no), Patrick Kück (patrick_kueck@web.de)

Objectives: Advances in high-throughput sequencing and genomics have revolutionized research in evolutionary biology and systematics. The use of genomics data in phylogenetic analyses has brought new challenges in terms of data handling and analysis. This course aims to help those that have basic experience in bioinformatics and molecular phylogenetics, and have projects focused on high-throughput sequencing data and phylogenetics, to become acquainted with tools, programs and pipelines for phylogenomics and want to contact phylogenomic studies beyond the standard also addressing potentially confunding biases in their datasets.

Prerequisites: Basic knowledge of command line interaction, regular expressions, UNIX, sequence data and alignments, model-based phylogenetic analyses, model testing, Bayesian inference, is a prerequisite for participation. Pre-course exercises on command line, regex and UNIX will need to be done, submitted and passed to ensure that all participants have an equal minimum level at start.

Travel and accommodation: The course is arranged by the Research School in Biosystematics – ForBio. There is no course fee for ForBio members and associates. Norwegian ForBio members will be covered for accommodation and refunded for travel expenses. ForBio will not refund travel or accommodation costs for other members and associates. See http://www.forbio.uio.no for more information on ForBio and membership.

Content

The workshop will consist of lectures and hands-on work aiming to provide the participants with a good working knowledge of genomic data for phylogenomics, and to become acquainted with tools, programs and pipelines for phylogenomics. To encourage discussion, participants will work in pairs most of the time. All exercises will be carried out using example data handed out by the teachers during the workshop, but all exercises and results can be readily transposed to individual research data.

Tentative schedule, week I, 14-18 Nov 2016

THS – Torsten Hugo Struck, PK – Patrick Kück

Mon am	Introduction & review of sequencing technologies (THS)
	Crash course working on the command line (the students will have to do some pre-course exercises to send in before the class, so that they are on a similar level)
	Sequence data formats and quality control measurement (THS)
Mon pm	Assembling genomes & transcriptomes with Ray & Trinity (THS & PK)
Tues am	Assessment of assembly quality (THS)
	BLAST at the command line (THS)
Tues pm	Annotation with Trinotate & MITOS (THS & PK)
Wed am	Orthology assignment with Orthograph (PK)
Wed pm	Sequence alignment with MAFFT and PAL2NAL (THS & PK)
	Masking with Aliscore and Alicut (PK)
Thur am	Assessment of orthology (PK)
	Concatenation with FASconCAT (PK)
	Matrix reduction with MARE (PK)
Thur pm	Tree reconstruction with RAxML and Supertree approaches (THS & PK)
	Model selection with RAxML (THS & PK)
Thur pm	Course dinner – social event
Fri am	Pipeline building (THS)
	KEGG & GO annotation (THS)
Fri pm	Summary & Wrap up (THS & PK)

Tentative schedule, week II

Week II will be taught in March or April 2017. We will provide the exact dates as soon as possible. The first week covers the basic foundation of phylogenomic studies. Lectures and labs in week II will build upon this and focus on dissecting the phylogenetic signal in the dataset and detecting possible misleading biases such as long branches and compositional heteroegeneities. Strategies for thorough sensitivity analyses in phylogenomic studies will be part of this week.