ForBio course: Phylogenomics,

Marine Biological Station Tollboden,
Drøbak, UiO June 03 – 15, 2018

About the workshop

Time: Jun 03 – 15, 2018.

Place: Marine Research Station Tollboden, Drøbak, University of Oslo

Course credits: 5 ECTS

Assessment: Written report on a provided problem during the course within two weeks after the closure of the workshop.

Registration: Registration before April 15th 2018

Fee: No fee for ForBio members or associates.

Maximum number of participants: 12

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 confounding 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. Moreover, we will provide a pre-course afternoon session on basic UNIX commands to refresh the knowledge, for accepted students asking for such a refresher.

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.

After covering the basic foundation of phylogenomic studies we will show how to dissect the phylogenetic signal in the dataset by focusing on different sources of incongruences and misleading biases. These are rogue taxa, base composition, branch length and evolutionary rate heterogeneity, saturation, paralogy, incomplete lineage sorting, horizontal gene transfer, ancestral hybridization, and contamination. Strategies for thorough sensitivity analyses in phylogenomic studies will be discussed during these weeks.

Throughout the course all exercises will be carried out using an example dataset, which will be build up during the course (starting data will be handed out by the teachers at the beginning of the course). During the course the evening hours are reserved for discussions about problems with phylogenomic data from the literature or with the student's own data as well as presentations of the students' projects, if they wish to. All exercises and results can be readily transposed to individual research data.