

2. Genome

Seminar Genome; Sequencing, Assembly & Annotation

Introduction

The seminar *Genome; Sequencing, Assembly & Annotation* will provide deeper insight in what a genome sequence comprises, and explain its dynamic nature. To obtain a genome sequence in general, the orders of base pairs are deduced from fragments of DNA following chemical or physical reactions (sequencing). The orders of the fragments are in turn linked using programs that calculate the most likely linkages based on statistics (assembly). Finally, patterns are searched in and recorded for the assembled fragments (annotation). Together this will result in an **approximation** of the **physical genetic map** of the sequenced individual and the **encoded information herein**.

It is crucial to understand that none of the steps during the generation of a genome sequence are ever without errors and a genome sequence is therefore not absolute, in contrast to what most people think. Also, species-specific properties such as genome complexity, or for example ploidy, limit the maximum possible genome sequence quality, and 'a genome sequence' can therefore differ substantially in information content between different species. Finally, the genome sequence of an individual, no matter how high in quality, is never *the* blueprint of a whole species. The individual of which genomic DNA has been extracted for sequencing will always possess some specific mutations or other unique properties in comparison with other individuals of the same species.

Obtaining a fully complete and error-free genome sequence is virtually impossible, and even a good approximation is often a complicated and costly endeavour. It is therefore wise to select an approach that optimizes those aspects of the genome sequence that are most important for the success of a given project. One should realize that obtaining a genome sequence by itself is seldom the goal, but instead is a mean to achieve a goal. In this seminar we will investigate the currently leading sequencing, assembly and annotation methods, examine their benefits and drawbacks, and explore possibilities for their combination.

Content

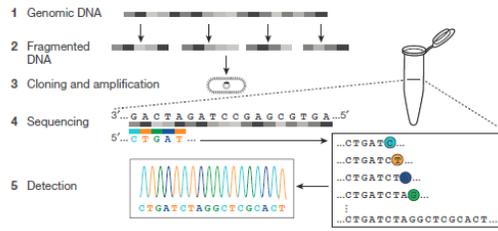
(1) Basics, performed individually. Reading of the review *A field guide to whole-genome sequencing, assembly and annotation, Ekblom et al., 2014*. <https://doi.org/10.1111/eva.12178> and answering of the accompanying general questions about genome sequencing, assembly and gene prediction. Some of the questions require a little extra information outside of the indicated review, which can be obtained by searching the internet (e.g. through Google).

(2) Specification, performed group wise. Students will groupwise study a selected set of sequencing, assembly or annotation methods, using per-group indicated reviews, websites and technical manuals. Students should be able to explain the key features of their set of techniques or methods and prepare a small presentation. Questions are provided for guidance on each group topic in addition to the suggested reviews, websites and manuals. Including the answers to these questions in the presentation will help to identify the key features.

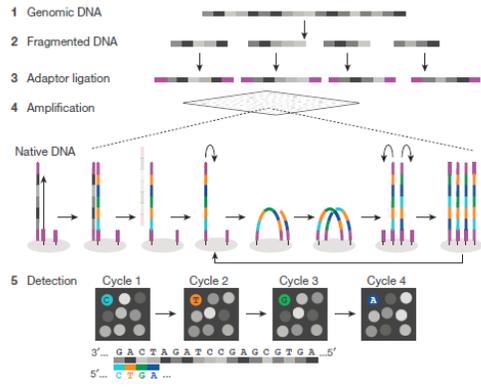
(3) Exchange, performed classically. Each group will present their group topic to the other students. Each presentation will be followed by questions from the other students, about details or unclarity on the presented techniques or methods or their possible uses. A good understanding of the different topics by all students will be needed to solve the presented casus. The casus is aimed at a sequence project in plants. Strengths and weaknesses of different methods and techniques will be compared and discussed during the casus.

DNA sequencing technologies. Click to enlarge.

First generation sequencing (Sanger)



Second generation sequencing (massively parallel)



Third generation sequencing (Real-time, single molecule)

