

Bioinformatics Practicals In Sillico
Project

Thibault Schowing
Lionel Rohner
Alain Rohrbasser
Rares Cristea

October 16, 2019

Yeast Genome Analysis

File types used

- **FA** The files with the .fa extension store FASTA format sequences. In this project the .fa file contains the reference genome.
- **GTF** The Gene transfer format (GTF) is a file format used to hold information about gene structure. It is a tab-delimited text format based on the general feature format (GFF), but contains some additional conventions specific to gene information. A significant feature of the GTF that can be validated: given a sequence and a GTF file, one can check that the format is correct. This significantly reduces problems with the interchange of data between groups.
- **VCF** The Variant Call Format stores the gene sequence variation. By using the variant call format only the variations need to be stored along with a reference genome which make the file less redundant.
- **BAM** Binary Alignment Map (BAM) is the comprehensive raw data of genome sequencing; it consists of the lossless, compressed binary representation of the Sequence Alignment Map. BAM is the compressed binary representation of SAM (Sequence Alignment Map). BAM is in compressed BGZF format.

Arabidopsis Thaliana Genome Analysis

Part 2

Lactobacillus Heleveticae Genome Analysis

Part 3