

## Exercise sheet 2

Group: Anders Busch, Dan Sebastian Thrane, Frederik Hertzum, Henrik Schulz, Lars Thomasen and Peter Gottlieb

### Explain/define: “Operon” and “Transcription Unit”?

- **An operon** is a functioning unit of genomic DNA containing a cluster of genes under the control of a single region of DNA that initiates transcription of a particular gene.
- **Transcription** is a process that occurs in all living cells. During transcription, strands of RNA are created based on the DNA found within the cells. When a strand of messenger RNA (mRNA) is created, it is then used to produce proteins during translation. A whole strand of DNA is not usually transcribed into mRNA, but instead specific sections of the DNA are, which are called transcription units.

**Write a Java program SELO that reads a file with the following eight sequences. SELO shall compute and output a table with the characteristic numbers for a “standard” sequence logo (no HMM logo).**

- See separate java folder.

**What is the main idea behind the unsupervised operon prediction method introduced in the class, i.e. why does it “work” without prior knowledge about concrete previously identified operons (sample/training data)?**

- TODO

**Why are the features for adjacent genes computed separately for closely and distantly related organisms?**

- TODO