Exercise set 2

1. The file `genomic\_dna\_3.txt` contains a DNA sequence that has two exons, separated by an intron. The first exon starts at the beginning of the sequence, and ends at the 63rd character, the second exon starts at the 91st character, and runs up to the end of the sequence. Read the file, and write the coding and non-coding regions to two files, `genomic\_dna\_3\_coding.txt` and `genomic\_dna\_3\_noncoding.txt` respectively.
2. Write a file `sequences\_3.fasta` in FASTA format for the following three sequences and their respective headers. Sequences must consist only of the symbols A, C, G, and T.  
   headers: 'ABC123', 'DEF456', 'HIJ789',  
   sequences: 'ATCGTACGATCGATCGATCGCTAGACGTATCG', 'actgatcgacgatcgatcgatcacgact', 'ACTGAC-ACTGT--ACTGTA----CATGTG'.
3. Using the data for the previous part, write three, rather than one file, i.e., a FASTA file per sequence. Use the sequence headers as the respective file names, adding the `.fasta` extension at the end.
4. The file `dna\_seqs\_4.txt` contains DNA sequences, one per line. Each sequence starts with the same 14 base pair fragment, a sequencing adapter that should be removed. Writa a program that reads the sequences from the input file, and writes the trimmed sequences to a new file `trimmed\_dna\_seqs\_4.txt`, and prints the length of the original and the new sequences to the screen.
5. Given a file `genomic\_dna\_4.txt` containing a DNA sequence, and a file `exons\_4.txt` with exon positions, one start and end position per line. Write a program that reads the positions, and splices out the introns in the given DNA sequence, writing the output to `genomic\_dna\_4\_coding.txt`.