1. EXERCISE (fasphy.pl)
Try converting the alignment stored in the simple FASTA file align.fas into a PHYLIP file called align.phy (there already is one on the website so you know how it should look like). Print your results to the screen first, so you can quickly check if your script works! HINT: There some nasty spaces trailing some lines in the FASTA file. Make sure to take care of them!

Description of the formats:

FASTA
- A leading ">" symbol indicates a line with the name of an individual
- The actual sequence of that individual is stored in the next line
- These pairs of lines are repeated for each individual in the alignment

PHYLIP
- The first line is header that contains the number of sequences and the length of the alignment separated by a space
- Each following line contains the name of the individual plus its sequence
- The name has to be exactly 10 characters long. If it is shorter it's filled up with spaces
- The sequence is given immediately after the name of the individual in the same line

2. EXERCISE (tcount.pl)
Open the file "map.sam" in Perl and loop over it using a "while(defined)" loop. Use this to count the number of hits for each transcript (identifiable by the FBtr number). Create an output file "tcount.txt" that contains the names of the transcripts and their respective number of hits in tab-separated format. HINT: As always when dealing with lists and counting things, a hash is the easiest way to do it.

3. EXERCISE (filter.pl)
Open the file "dmel.gff" in Perl and loop over it using a "while(defined)" loop. Use this to extract information about the mRNAs in the file. The last column of the file contains information about the name (ID=FBtr...) and the corresponding gene (Parent=FBgn...) of each mRNA. Create a file called "t_to_g.txt" that contains the name of each transcript and its parent gene in tab-separated format. HINT: While parsing through the file line-by-line, use the "split" function and pattern matching to extract the information you want.

4. EXERCISE (gcount.pl)
Use the information in "tcount.txt" and "t_to_g.txt" to add up all hits on transcripts belonging to the same gene. Create a file called "gcount.txt" that contains the name of each gene and the sum of all hits on transcripts belonging to that gene. HINT: You will have to create a Transcript -> Gene "database" coming from the "t_to_g.txt" file first, perfect for a hash! Also, the adding up of reads per gene is most easily done in a hash...