Week 3 – Bioinformatics File Formats Assignment

1. 1. What type of data is stored in a FASTA file?

A FASTA file stores DNA, RNA, or protein sequences. Each entry starts with a name or label line (which begins with a '>'), followed by the actual sequence written in letters (A, T, C, G for DNA, or amino acids for proteins).

1. 2. Why do we use BAM instead of SAM?

BAM is a smaller, compressed version of SAM, so it takes up less space on the computer. It’s also faster to read and works better with most tools, especially when dealing with large data.

1. 3. What do REF and ALT mean in a VCF file?

In a VCF file, REF is the normal or reference version of a DNA sequence at a certain spot. ALT is the alternative version found in the sample being tested — the difference or change from the reference.