First, you need to identify which bioinformatics file format we're dealing with. Common formats include FASTA, FASTQ, SAM/BAM, GFF, BED, and VCF.

1. FASTA/FASTQ file:

Since FASTA and FASTQ have simple structure, we can add the Gene Ontology to the header line:

>sequence\_identifier | GO:0006915 | Additional description or metadata

sequence\_data

some example code: add\_ontology\_to\_fasta.py

If we have a large number of sequences and need to map many of them to ontology terms, it might be more efficient to preload a dictionary of sequence identifiers to ontology terms rather than making a lookup for each sequence.(related to Ontology function)

1. SAM(https://samtools.github.io/hts-specs/SAMv1.pdf)

SAM (Sequence Alignment/Map) Headers contain @CO lines, which are free-form comment lines. We could use these lines to store ontology data relevant to the entire dataset or experiment.

If our ontology data pertains to individual reads/alignments, we’ll need to consider the TAGs in the alignment section. SAM format supports optional fields in the alignment section where we can define custom tags.

1. BAM

BAM need to convert to SAM and then use the function of SAM

1. GFF
   1. GTF

* 1. GFF3