Glossary of less known terms that are used in this course

- **Adaptors:** a short chemically synthesised single or double-strand DNA (oligonucleotide) used in some sequencing methodologies that allow the addition of a DNA barcode or other oligonucleotides downstream of an unknown or amplified DNA strand.

- **Amplicon:** a DNA fragment that has been amplified using polymerase chain reaction (PCR) or a method that results in the generation of multiple copies of that fragment.

- **A-tailing:** an enzymatic reaction to add a sequence of adenines at the 3'-terminus of a DNA fragment for sequencing purposes.

- **BAM:** Binary version of a SAM file format.

- **Barcoding:** addition of a tag of known DNA sequence (barcode) to an amplified DNA strand that permits sequencing multiple samples in parallel, and stratifying sample data informatically post/during sequencing.

- **cDNA:** complementary DNA, a DNA molecule synthesised from an RNA molecule.

- **COVID-19:** coronavirus disease 2019.

- **DNA:** deoxyribonucleic acid is an information molecule forming the "base code" for a living organism.

- **Enzyme:** a protein able to catalyse, i.e. accelerate chemical reactions.

- **Exome:** part of genome composed of exons, i.e. the sequences that will be transcribed and translated into proteins.

- **FASTA:** text-based format for representing nucleotide or peptide sequences, with base pairs or amino acids represented by single-letter codes.

- **PCR:** polymerase chain reaction (PCR) is a laboratory technique for rapidly making (amplifying) millions to billions of

copies of a given section of DNA, which may then be analysed further.

- **Primer:** a short chain of oligonucleotides used to target and 'prime' the initiation of DNA replication.

- **Protein translation:** is the process of synthesising proteins from a transcript.

- **NGS:** next-generation sequencing, a high throughput sequencing methodology.

- **Nucleotides:** the individual subunits that form DNA and RNA molecules.

- **RNA:** ribonucleic acid, an information molecule, can be the "base code" for viruses.

- **SAM:** sequence alignment/map format. It is a TAB-delimited text format that holds information from sequence alignment.

- **SARS-CoV-2:** severe acute respiratory syndrome coronavirus 2.

- **SNP:** single-nucleotide polymorphism is a nucleotide substitution in a specific position of the genome. It is identified when compared to a reference sequence.

- **Sonication:** is a technique that applies sound energy to a sample at a specific amplitude. In molecular biology, it is commonly used for fragment DNA fragmentation.

- **Transcription:** is the process of copying a segment of DNA into RNA.

- **Trimming:** in bioinformatics, it is the process of removing the ends of the reads, leaving only a region of high-quality bases and those relating to the sample, not the sequencing chemistry or DNA barcoding.

- **VCF:** variant call format, is the plain text file format that holds gene sequence variations.

- **Viral replication:** is the mechanism in how viruses propagate during the infection cycle, it is the process of virus multiplication.

- **Viral genome variant:** a virus that has one or more mutations in its genome.

- **Whole-genome sequencing (WGS):** is the method of determining the entirety, or almost the entirety, of an organism's genomic DNA sequence all at once. This entails sequencing all of an organism's chromosomal DNA, as well as, DNA found in mitochondria and, in plants, chloroplasts.