

Molecular Biology Basics

This is a compilation of resources to understand the biological basis of DNA sequencing and analysis. It explores what is DNA, how we amplify it in the lab to make it analyzable and what file formats and algorithms are used for this analysis.

Biology

- **DNA** (understand the basics of what is DNA: 4 different bases, 2 complementary strands, what is a mutation...)
- **Gene** (just to know what is a gene in general, GenBank database)
 - <https://my.clevelandclinic.org/health/body/23064-dna-genes--chromosomes>
 - <https://www.ncbi.nlm.nih.gov/genbank/>
- **DNA primer**
 - <https://www.genome.gov/genetics-glossary/Primer>
 - [https://en.wikipedia.org/wiki/Primer_\(molecular_biology\)](https://en.wikipedia.org/wiki/Primer_(molecular_biology))
- **PCR aka Polymerase Chain Reaction** (primer, amplification, ...)
 - <https://www.youtube.com/watch?v=wBrNbbAIAFo>
- **Sequencing** (what is the meaning of DNA sequencing)
- **Oxford Nanopore MinION** -> This is the device we used for sequencing at GenoRobotics, you can check what it is but you don't absolutely need to understand the exact biological mechanism behind it (but it's still useful to know if you are interested)
 - <https://nanoporetech.com/products/minion>
 - https://www.youtube.com/watch?v=1_mER5qmaVk
 - <https://www.youtube.com/watch?v=RcP85JHlMnI>
- **DNA barcoding** (general concept, it's the method we are using to identify different species)
 - https://en.wikipedia.org/wiki/DNA_barcoding
 - <https://www.youtube.com/watch?v=HpRHhIbDFd8>
- For the biological concepts, you don't need to understand everything in detail, you just need to know the basics/concepts and the things I mentioned.

INFORMATICS

- **.fasta and .fastq** file formats
 - https://www.youtube.com/watch?v=cJm_BGpinWg

- <https://www.ncbi.nlm.nih.gov/genbank/fastaformat/>
- <https://zhanggroup.org/FASTA/>
- **Sequence alignment algorithms**
 - **Needleman-Wunsch algorithm** (global alignment, match/mismatch/gap scores)
 - https://en.wikipedia.org/wiki/Needleman%E2%80%93Wunsch_algorithm
 - <https://www.youtube.com/watch?v=ipp-pNRlp4g>
 - **Smith-Waterman algorithm** (local alignment, kind of the same principle as Needleman-Wunsch)
- **"biopython" Python library** (you don't need to read it in detail, for the moment you just need to know that it exists because we will use it a lot)
 - <https://biopython.org/wiki/Documentation>

For the informatic part, it's very important to understand the file formats fasta/fastq and the Needleman-Wunsch algorithm. You can obviously use other resources/websites to learn these concepts (in English or French) and I just put some links as an example.