

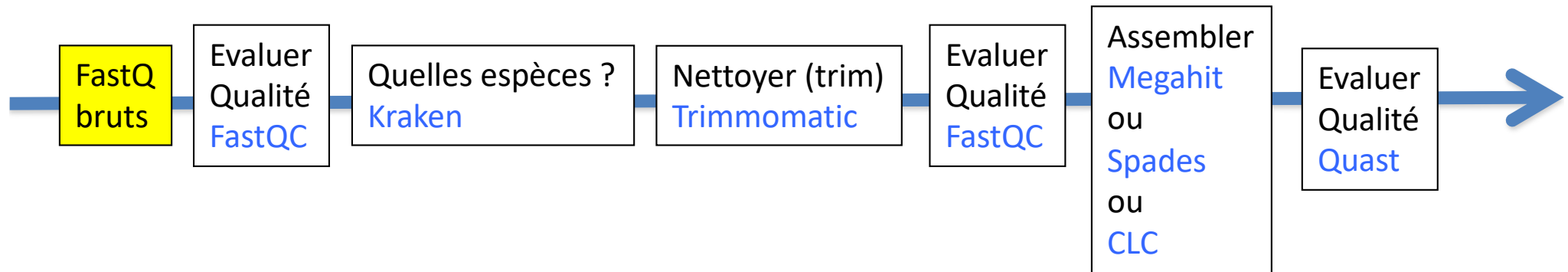
Ils sont compressés:

SampleName_S1_L001_R1_001.fastq.gz

SampleName_S1_L001_R2_001.fastq.gz

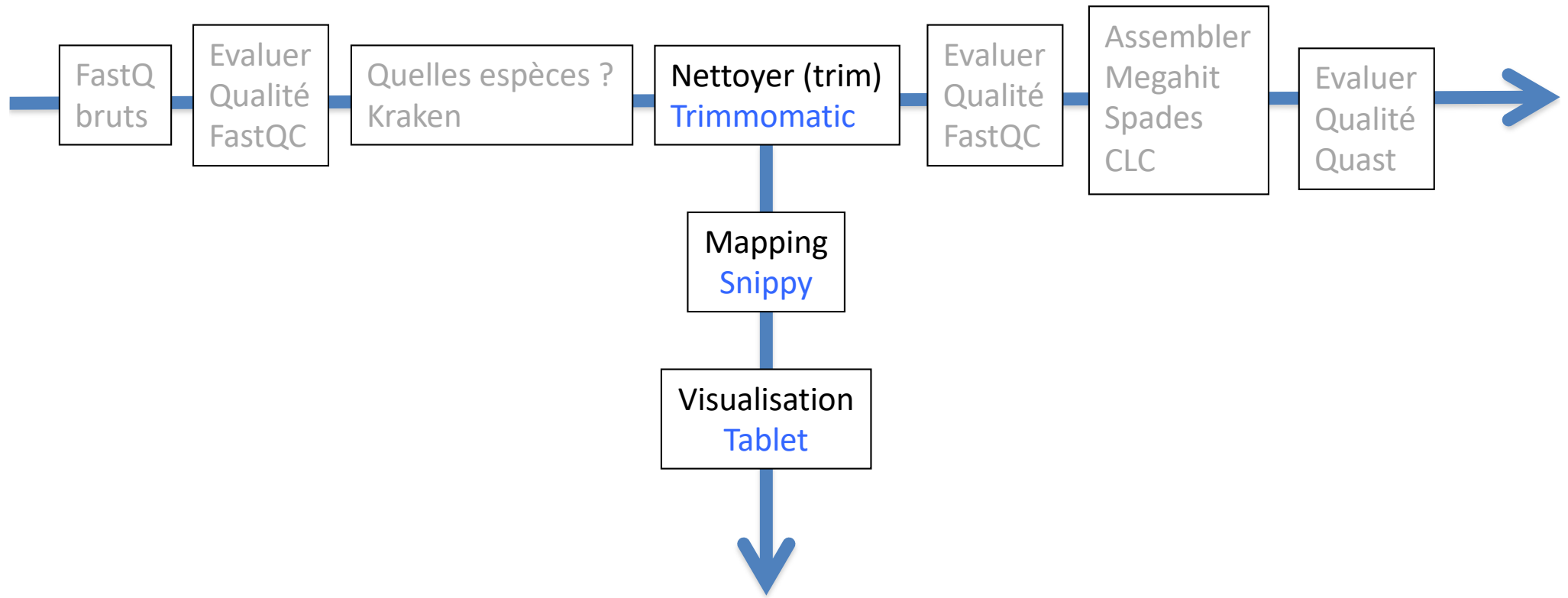
ILLUMINA Naming Convention

- ▶ SampleName—The sample name provided in the sample sheet.
- ▶ S1—The sample number based on the order that samples are listed in the sample sheet starting with 1.
- ▶ L001—The lane number.
- ▶ R1—The read. For a paired-end run, there is at least one file with R1 and R2.
- ▶ 001—The last segment is always 001.



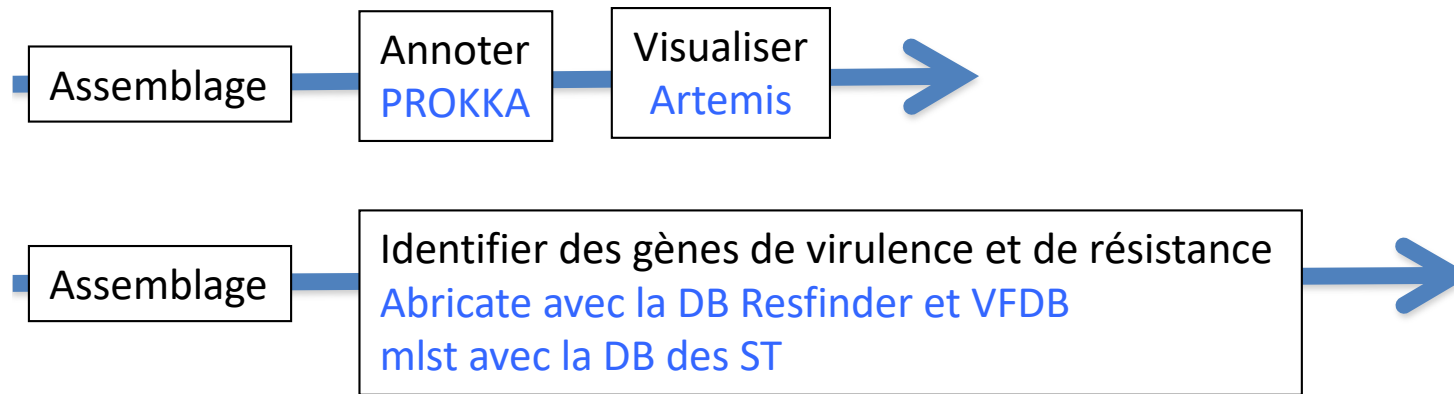
WORFLOW n° 1

- À partir de fichiers bruts de séquences (.fastq.gz)
- Nous allons évaluer leur qualité et regarder ce qu'elles contiennent
- Nous allons les nettoyer puis vérifier les conséquences du nettoyage
- Nous allons assembler les séquences nettoyées pour obtenir des « contigs » (.fasta ou .fa ou .fna ...)
- Nous allons comparer plusieurs assemblages différents pour choisir le meilleur



WORFLOW n° 2

- À partir de fastq nettoyés
- Nous allons mapper ces reads contre un génome de référence
- Nous allons visualiser le mapping et les SNPs et chercher des régions du génome de référence non couvertes



WORKFLOW n° 3

- À partir des assemblages
- Nous allons chercher les ORFs et leur attribuer une fonction = annotation du génome
- Nous allons visualiser notre fichier d'annotation (.gff ou .gbk)
- On peut aussi ne rechercher que les gènes présents dans une DB (VFDB, Resfinder, MLST...)

