Box GI.2. High-throughput sequencing methods.

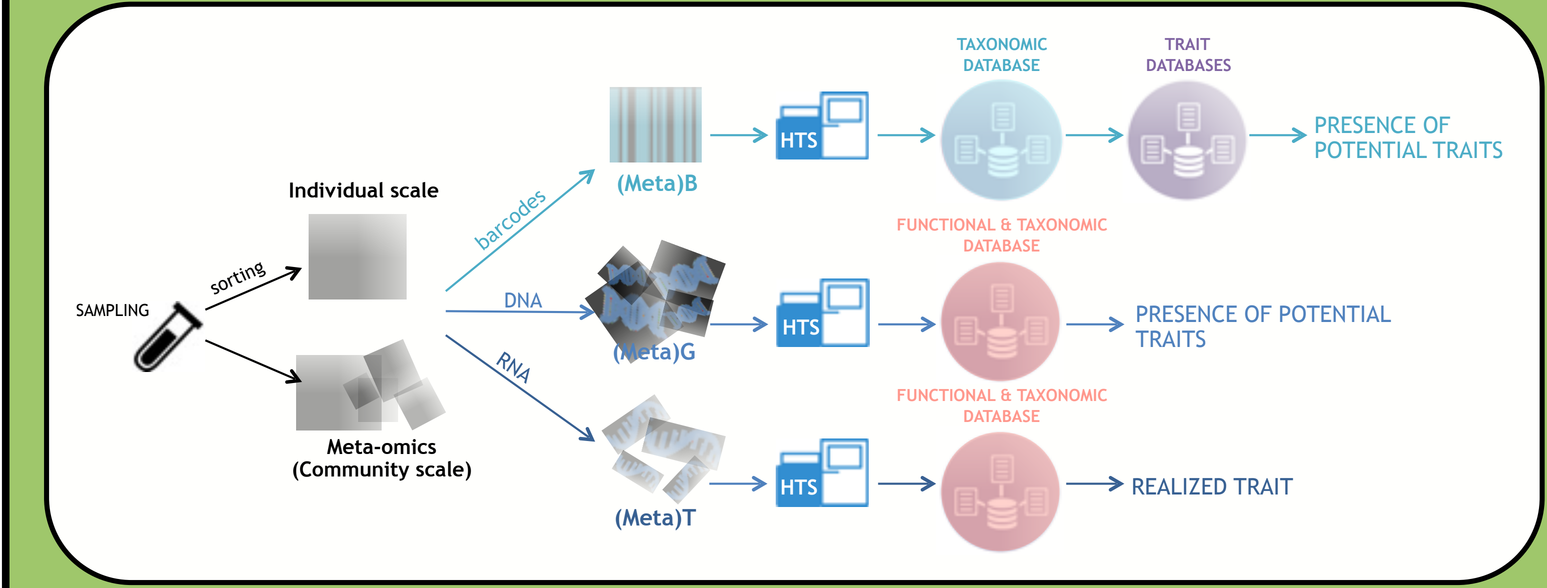

Figure GI.5. Most common High-Throughput Sequencing methods in microbial ecology.

High-Throughput Sequencing (HTS) methods are techniques used to sequence DNA or RNA that appeared in the 1990s and that have been democratised in the 2000s. Compared to the older Sanger sequencing (Sanger *et al.*, 1977), HTS produces more sequences, at faster speeds and cheaper costs. In 2024, sequencing data used to investigate microbial communities usually come from 2nd HTS generation techniques (that appeared in the mid-2000s), mostly from Illumina (Abdelfattah *et al.*, 2018). They produce short reads compared to Sanger sequencing (*e.g.*, Illumina: from 50 to 500 bp, Pyrosequencing 454: from 300 to 600 bp). 3rd HTS generation techniques have been developed since ~ 2015 (*e.g.*, PacBio, Nanopore) (Giordano *et al.*, 2017). They produce longer reads (> 1000 bp), but their costs have prohibited them from being widely used yet.

Three sequencing methods are widely used in microbial ecology: **metabarcoding** (metaB), **metagenomics** (metaG) and **metatranscriptomics** (metaT). They are adaptation to the whole community of their single-organism counterparts barcoding, genomics and transcriptomics (Figure GI.5).

All these methods start from the extraction and purification of DNA or RNA from a population of the same organism or of a whole community in the case of meta-methods (*e.g.*, obtained by filtering water through a polycarbonate filter).

In the case of **metaG**, the whole DNA is sequenced. In the case of **metaT**, RNA is reverse-transcribed to complementary DNA (cDNA), which is then sequenced. In both cases, short reads are usually assembled into longer sequences termed contigs based on their overlapping sections using dedicated programmes (*e.g.*, MEGAHIT (Li *et al.*, 2015)). Those contigs in turn are given a taxonomy (*e.g.*, using BLAST (Camacho *et al.*, 2009)) and a function (*e.g.*, using InterProScan (Jones *et al.*, 2014)) by comparison to reference sequences in databases (*e.g.*, RefSeq for taxonomy, Pfam, PANTHER, PRINTS for function). As DNA is highly stable and RNA exhibits a fast turnover, **MetaG** gives information on potential traits the organisms may exhibit, while **metaT** informs on which genes are actually transcribed, which gives a better proxy of which traits the organisms may exhibit (Aguiar-Pulido *et al.*, 2016). Recent methods based on k-mers (*e.g.*, Kraken2 (Wood *et al.*, 2019)) allow to skip the assembly step and to directly give taxonomic annotations to short reads.