- De novo assembly is a method for constructing genomes from a large number of (short- or long-) DNA fragments, with no a priori knowledge of the correct sequence or order of those fragments.
- The goal of a sequence assembler is to produce long contiguous pieces of sequence (contigs) from these reads. The contigs are sometimes then ordered and oriented in relation to one another to form scaffolds.
- There are two types of algorithms that are commonly utilized by these assemblers: greedy, which aim for local optima, and graph method algorithms, which aim for global optima. Different assemblers are tailored for particular needs, such as the assembly of (small) bacterial genomes, (large) eukaryotic genomes, or transcriptomes.
- We propose a new method to correct short reads using de Bruijn graphs and implement it as a tool called Bcool
- tools following the de Bruijn graph (DBG) paradigm generally attempt to filter out erroneous k-mers by considering only k-mers present at least a minimal number of times in the reads to be assembled. Both paradigms may benefit from a preliminary error-correction step