reaBuilding a pipeline for analysis of short-d RNA-seq data and natural transcriptomic variation in Caenorhabditis elegans

Cristian Riccio, Mélanie Tanguy, Lise Frézal

C. elegans is a model organism that has been used extensively since its establishment by Sydney Brenner in 1974. 60-80% of human genes have an ortholog in C. elegans. Hence, many findings in C. elegans are transposable to humans. However, the laboratory strain N2 that was isolated in Bristol in the fifties has diverged extensively from its wild-type counterparts. As shown in this and previous studies, there is great genetic diversity in C. elegans. The genetic diversity, combined with phenotypic diversity, enables us to study the genetic component of phenotypic traits. Using C. elegans as a model organism for the study of genome-phenotype associations has many advantages over the study of humans: higher throughput due to lower experimental costs, controlled environment, viewed as ethically less problematic by many. In this study, we have sequenced the non-ribosomal RNAs of 12 strains of C. elegans to study the natural variability of gene expression in the nematode. We have built a computational pipeline that performs quality assessment and filtering, alignment, transcript counting and variant calling. Furthermore, we have identifed a total of 43769 nucleotide variants across the 12 strains, suggesting that genetic variability in wild isolates of C. elegans is not negligible. This piece of work is the first step towards understanding the role of the genetic and environmental backgrounds in the determination of the phenotype of C. elegans.