# Prediction of sequence divergence from the quality of mapping

Kovacova Viera1, Lässig Michael1

1 IBP, University of Cologne, Zülpicher Str. 77, Cologne, Germany

Read mapping, one of the first steps in sequencing data processing, can significantly influence downstream analysis results. Simulated sequencing libraries with different levels of sequence divergence mapped to the same reference genome display drop-off of mapped reads from tenths to tens of per cent. For example, if the sequence divergence is close to seven per cent, we see a considerable decrease of mapped reads under the default setting of the STAR aligner.

Correctly adjusting a reference genome and testing the mapping tool's parameters to decrease false findings is crucial, mainly in microbial studies. On the other hand, known proportions of unmapped reads under defined mapping parameters can inform us about sequencing divergence and evolution of given (sub)species.

To compare the results of simulated data sets with natural sequences, we used B. subtilis's subspecies spizizenii and subtilis, where speciation led to nearly 7% sequence divergence.