

# SeqTrim for pre-processing in the next-generation sequencing era

M. Gonzalo Claros<sup>1</sup>, N. Fernández-Pozo, D. Guerrero, R. Bautista, A. Bocinos, J. Falgueras

## Abstract

High-throughput automated sequencing has enabled an exponential growth rate of sequencing data. This requires increasing sequence quality and reliability in order to avoid database contamination with artefactual sequences and to rely on pre-processing results without further visual inspection. SeqTrim was designed to gather these capabilities.

As a standalone and Web application, it seems to outperform other pre-processors in the identification of the true insert sequence devoid of low quality, vector, adaptor, low complexity and contaminant sequences. Changing appropriately its parameters, SeqTrim performs equally well with sequences from EST libraries, SSH libraries, genomic DNA libraries and pyrosequencing reads, and does not lead to over-trimming. Its friendly interface allows users to know what happened with sequences at every pre-processing stage, and to verify pre-processing of an individual sequence if desired. A detailed description of every modification undergone by the sequence is given.

A practical hand-on session will permit participants to know how to use SeqTrim Web interface with different kind of sequencing experiments, to realise the undesirable effects of a mis-pre-processed sequence, and to obtain high-quality reads from a pyrosequencing run. SeqTrim can be freely accessed (and downloaded) from <http://www.scbi.uma.es/seqtrim>.

---

<sup>1</sup> Plataforma Andaluza de Bioinformática, Universidad de Málaga, Málaga, Spain