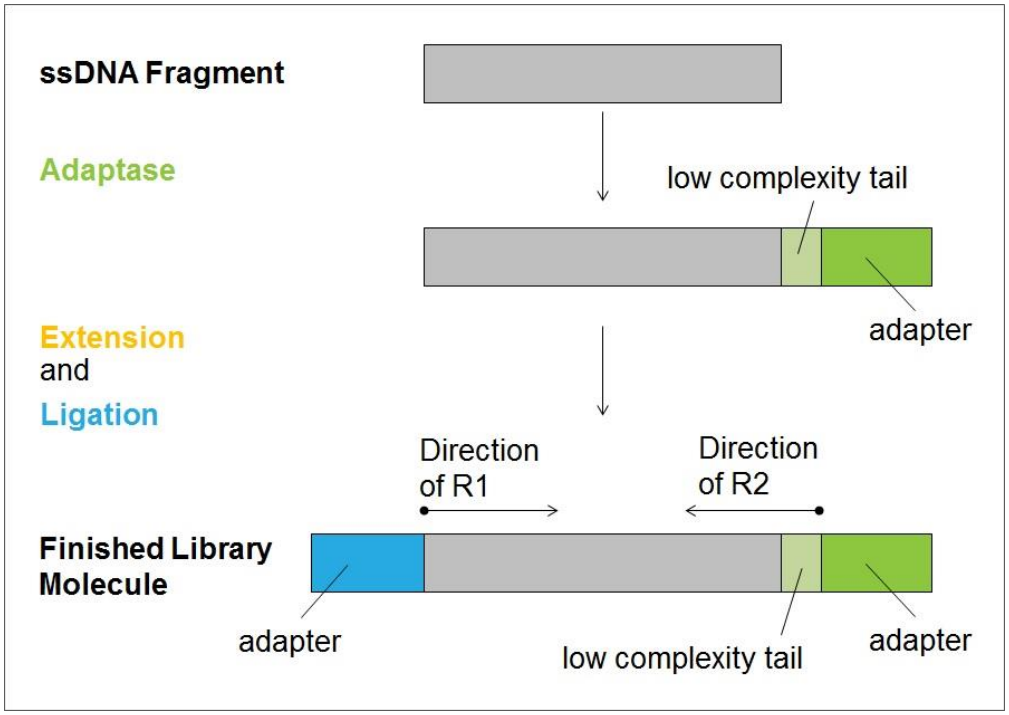
ACCEL-NGS® 1S PLUS & METHYL-SEQ: TAIL TRIMMING FOR BETTER DATA

INTRODUCTION

Swift Biosciences' Adaptase™ technology, used in the Accel-NGS 1S Plus and Methyl-Seq kits, adds a low complexity polynucleotide tail with an average length of 8 bases to the 3' end of each fragment during the addition of the first NGS adapter molecule. If these tails are not trimmed bioinformatically from the sequencing data, it is normal and expected to observe them at the beginning of Read 2 (R2). When read length is close to fragment size, the tail may also be observed toward the end of Read 1 (R1) data. This technical note describes the nature of these tails and provides guidance for when and how they should be trimmed.

Figure 1: The Adaptase Step



The Adaptase Step is a highly efficient, proprietary reaction that simultaneously performs end repair, tailing of 3' ends, and ligation of the first adapter to 3' ends. The extension and ligation reactions finish the library molecule by adding the second adapter to 5' ends.

SEQUENCING RECOMMENDATIONS

Illumina® sequencing chemistry, which uses the initial bases of each read to perform cluster registration and establish focus and color balance, is sensitive to low complexity base composition at the start of the read. Therefore, it is important to sequence Accel-NGS 1S Plus and Methyl-Seq libraries on the MiSeq® using MiSeq Software Updater v2.2.0.2 (containing RTA v.1.17.28 rel. 3/18/2013) or later; or, if using the HiSeq®, assign a control lane loaded with a balanced, high-complexity sample. If using the HiSeq in Rapid Run mode, adding a high complexity spike-in is recommended. These precautions consistently lead to highly successful sequencing runs with Q-scores above 90%. Quality control software, such as FastQC (Babraham Bioinformatics), may raise "Per base sequence content" or "Per base GC content" flags at the beginning of R2. These flags are expected due to the low complexity tail.

WHEN TO TRIM: DNA LIBRARIES MADE WITH ACCEL-NGS 1S PLUS

For whole genome or other DNA sequencing applications (like ChIP-seq) involving alignment to reference genomes using common aligners, such as BWA-MEM, BWA-ALN (Li H. and Durbin R., 2009), or Bowtie 2 (Langmead B. and Salzberg S., 2012) or de-novo assemblies, trimming of the adaptase tails is recommended. Depending on the aligner and parameters used, tail trimming may significantly improve mapping efficiency. Therefore, standard adapter trimming should be followed with tail trimming (see below).

WHEN TO TRIM: ACCEL-NGS METHYL-SEQ LIBRARIES

The Accel-NGS Methyl-Seq kit adds bases to 3' termini during the Adaptase tailing step, including unmethylated cytosines. This tail adds both artifactual sequence and methylation information to the dataset. Therefore, trimming is **required** for Accel-NGS Methyl-Seq libraries to obtain improved mapping efficiency (with tools like Bismark or BSMAP) and precise methylation information.

Many informatics pipelines for Methyl-Seq analysis already include trimming of up to 10 bases from the beginning of both R1 and R2 to eliminate any artifactual cytosine methylation introduced as a result of filling in overhangs during end repair steps of conventional dsDNA library preparation and low quality bases due to bisulfite treatment. See below for ideal trimming recommendations.

HOW TO TRIM

Paired-End Sequencing

Trimming **10 bases from the beginning of both R1 and R2** following adapter trimming eliminates the majority of Adaptase tails. Publicly available tools like Trimmomatic (Bolger, et al. 2014 Bioinformatics) or Cutadapt (Martin, et al. 2011 EMBnet.journal) may be used to perform this trimming as part of the sequence data processing pipeline. This is necessary to maximize mapping efficiency for Adaptase-based libraries and to achieve accurate methylation analysis for Methyl-Seq. Although the majority of tails are at the beginning of R2, for paired-end alignment, it is recommended that R1 and R2 be trimmed symmetrically for better aligner performance and capability. Tail trimming must be performed following adapter trimming to ensure that both types of sequences are efficiently removed. When read length is greater than 100 bases, which increases the frequency of sequencing through short inserts, tail sequences will also appear at the end of R1. In this case, additional trimming of 10 bases from the ends of both R1 and R2 is recommended in order to remove the residual tail sequences encountered.

Single-End Sequencing

For read lengths greater than 100 bases, trimming **10 bases from the end of R1** following adapter trimming eliminates the majority of Adaptase tails. Publicly available tools like Cutadapt, Trimmomatic, or fastx_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/) may be used as part of the sequence data processing pipeline. This is necessary to maximize mapping efficiency for Adaptase-based libraries and to achieve accurate methylation analysis for Methyl-Seq. For read lengths less than 100 bases, tail trimming is not required.

For additional tail trimming recommendations, please contact Swift Biosciences' Technical Support at technicalsupport@swiftbiosci.com.

Swift Biosciences, Inc.
58 Parkland Plaza, Suite 100 • Ann Arbor, MI 48103 • 734.330.2568 • www.swiftbiosci.com