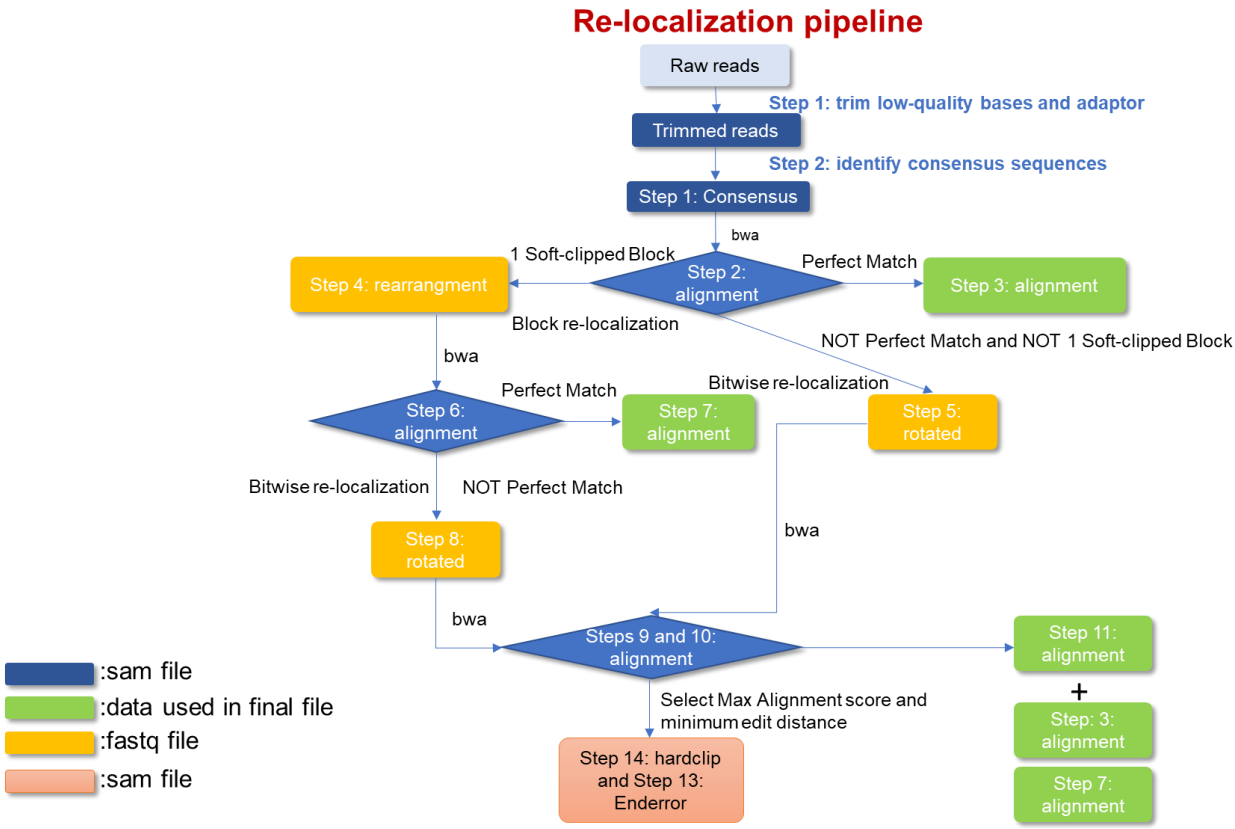

Relocalization pipeline. (Step 1): Raw read are trimmed with trimmomatic (Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.) (Step 2): Consensus are generated by a script modifying from file downloaded at https://andino.ucsf.edu/CirSeq. Modification are made to deal with irregular read length and overhang due to trimming. To achieve this, code had been modified from hard-coded to soft-coded. (Step 3): Consensus(CS) are alignment by bwa-mem instead of bowtie2 as bwa-mem performs better than bowties2. (Step 4): All reads that are perfectly matched with the genome print to a file called 3_alignment.sam.gz/7_alignment which will later combined to a file called data.sam.gz, script had been modified to fit the output format of bwa-mem. (Step 5): All reads that are not perfect match and NOT containing only one S block in cigar string are rotated to generate a combination of all possible sequence in fastq.gz format. (Step 6): All reads that had only one S block in CIGAR string are rearranged as original script did. (Step 7): All reads that contains only M block in CIGAR string with minimum edit distance(first priority) and maximum alignment scores(second priority) among the same sequencing ID will be printed to 11_alignment.sam.gz. (Step 8): Reads with minimum edit distance(first priority) and maximum alignment scores(second priority) among the same sequencing ID, but contain S or H block in CIGAR string, are printed to 13_Enderror.sam.gz(contain S blocks) or 14_hardclipped.sam.gz(contains hardclips)