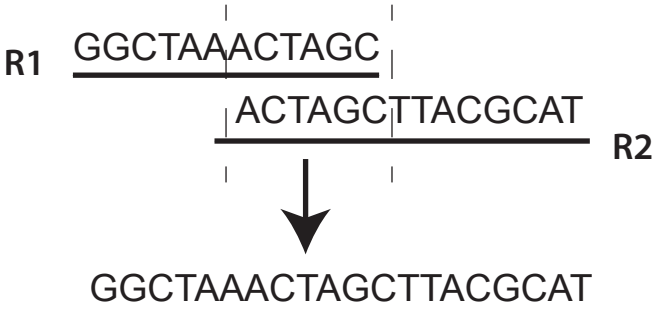
A Brief Guide to What Mothur is Doing (in the SOP)

From my understanding of Mothur, the stepts can be divided into two types, computational and operational. Computational steps are not relevant to the biology, but speed things up. These are steps like unique.seqs, pre.cluster, and the alike. The operational steps are relevant to the biology and chemistry of the sequences. Below is a brief summary of what Mothur is doing, operationally. There is no substitute for understanding both (https://www.mothur.org/wiki/MiSeq_SOP).

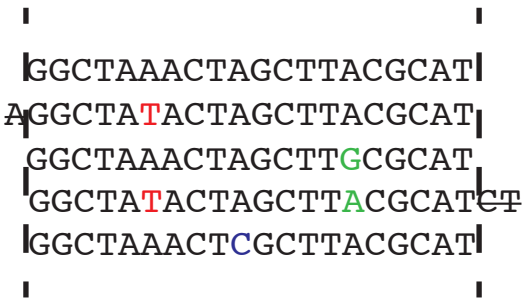
Step 1: Quality Control (QC) and Contigs

A) Get rid of sequences with bad quality.
B) Join paired-end reads (R1 & R2) into a single contig.



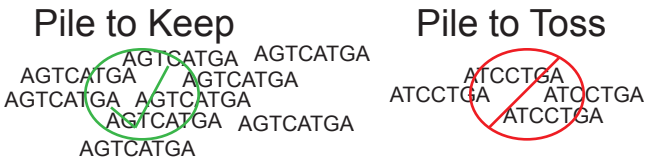
Step 2: Alignment

C) Align sequences using conserved positions
D) Get rid of positions outside of predicted sequence (i.e., areas outside of targeted amplicon)



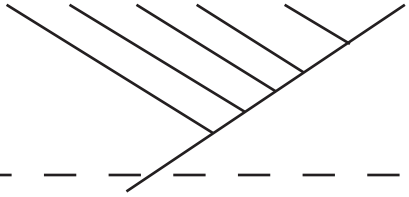
Step 3: More QC

E) Get rid of chimeric sequences caused by PCR.
F) Get rid of sequences that you are not interested in or were erroneously amplified.



Step 4: Cluster and Assign Taxonomy

G) Group sequences by similarity.
H) Assign them to taxonomic groups.



Step 5: Community Matrix

I) Generate a table with OTU x Sample.
J) Ready for downstream analysis.