

Strategies for Finishing (Hybrid Assemblies)

For easy use, this protocol is also available as a single sheet document on the GEP website. Most of the goals of the project (steps 1 through 4 below) can be attempted in any order, the following order is advisable if you will be doing your own PCR/Sanger reactions. The order is designed to allow as much time as possible for generating the PCR/Sanger data.

1. Use search for string to search for “nnnn”; this will generate a navigator listing the locations of any gaps found in the project.
 - a. If gaps are present, assess if an overlap exists; if an overlap is present, use tear/join technique (see demo in walkthrough) to remove gap.
 - b. If no overlap can be found; design PCR primers flanking the gap. Optionally, order primer synthesis and attempt PCR/Sanger in lab and incorporate any new data into the project.
2. Confirm that the consensus quality threshold is set to 30: Main window -> Options -> General preferences. Now, in the aligned reads window, use Navigate -> Low consensus quality menu to list all LCQ regions. Any LCQ region in the first 2.5 kb, the last 2.5 kb, or associated with a gap does not need to be inspected.
 - a. Inspect all appropriate LCQ regions, for regions with a consensus quality below 25 (or below 30 if the region is single-stranded); these regions should be manually tagged with a “Data needed”. Optionally, design and order primers and attempt PCR/Sanger in lab and incorporate any new data into the project (be sure to ask your mentor if you are doing the optional primer design or wet-bench work on the project).
3. Use the Main Window -> Navigate -> “Search for Highly Discrepant Positions” to generate the list of potential problem areas with at least 3 discrepancies with Q scores ≥ 30 .
 - a. Go to every region on the list with at least 3 HQD’s that is within 5 bases of a MNR that is at least 5 bases in length. Carefully examine the region to confirm or correct the length of the MNR in the consensus. If the region is not associated with a MNR, it can be ignored.
 - b. (Optional goal) Examine any region where the proportion of the HQD’s is between 40-60%. If the region is not marked with a repeat tag assess the likelihood of polymorphism vs. mis-mapping, and if appropriate, add a polymorphism tag to the location.
4. Use “Main Window -> Navigate -> Search for High (low) Depth of Coverage” to find all regions of score 10 with a 40 fold or less coverage. Carefully double check **all** MNR runs of length 5 or more found within these regions of low coverage. Minimum standard is 2 Illumina reads with all bases Q 20 in the MNR in order to overrule the consensus. Illumina reads with more than one HQD at any location not associated with the MNR should **NOT** be considered as evidence to support the consensus. This is to avoid introducing consensus errors from mis-mapped reads.
5. Be sure to complete and include in your submission the finishing report form. This is especially important if you improved the consensus but were not able to fully finish the project.