

PCR/Sanger pipeline and primer design

The PCR/Sanger pipeline discussed below can be used for adding Sanger data for any region. For simplicity resolution of a gap will be assumed.

Sanger sequencing reads of PCR amplicon can be incorporated into the assembly using the “Add New Reads” functionality in Consed. Because “Add New Reads” will only align the new reads against the existing assembly, finishers will need to force join any overlapping regions created by the addition of the Sanger reads to resolve gaps. Large gaps could require multiple rounds of sequencing of the PCR template before creating sufficient sequence overlap to justify manual joining. Given that we are working with areas that for some reason failed in the whole genome sequencing, it is probably not surprising that we find a higher failure rate of PCR amplification. This is particularly true for repeat rich regions. To improve the odds of success, finishers may wish to design **two sets of primer pairs** (as discussed below). Ideally both forward primers will be compatible with both reverse primers. This will allow for quickly testing 4 different PCR primer pairs for amplification. This recommendation, however, does increase the cost; the balance between time spent on task and the cost of supplies is left to the individual finisher in consultation with their mentor.

Should all attempts to resolve a region fail, a comment tag should be added to the region giving details of which primer pairs were tried. This will keep others who may subsequently work on the project from duplicating effort.

Also since every genome is different, everyone is encouraged to experiment to find ways of improving the success rate of the PCR amplification. Please share your findings with the group through the wiki or the bulletin board.

PCR/Sanger primer selection

Most projects will have a very small number of regions that need PCR/Sanger to resolve the NMR or gaps. Preliminary work suggests that half to a third of all projects will require no PCR/sanger sequencing; these projects simply need manual inspection and edits to correct consensus errors. For those rare regions that do need Sanger sequencing, you will want to take care to select the best primer pair. Sometimes there will be so few available primers that the choice of which primers to order will be trivial. However sometimes you will have many possible primer pairs. In these complex situations, expect to spend at least 15 minutes to select your primer pair. This will not guarantee successful PCR amplification and Sanger sequencing but it will improve the probability that the reactions will be successful.

If you are instructed to try PCR/Sanger one primer pair at a time, the choice of primers will be fairly straightforward. If you are instructed to pick 2 primer pair, try to pick 4 primers

that will all be mutually compatible. Try to keep the size of the PCR amplicon relatively small, ideally less than 1 kb. This will not always be possible especially in repeat rich regions, and it may not be advisable for regions with a cluster of issues. However, all other things being equal, it is a generally easier to amplifying smaller products than larger ones. Hence the recommendation to keep the size to less than 1kb. For sites with many possible primer pairs use the discussion below and the example in the hybrid assembly walkthrough to pick the best possible primers.

Primer selection and optimization

What we would like to do is pick primers that are unlikely to prime DNA synthesis at sites other than the region we wish to sequence (this phenomenon is often called “off-target” priming). To avoid off-target priming we will use BLAST to search the project genome for regions that might also prime DNA synthesis in the PCR reaction (see the recommended BLAST search parameters below). Almost all primers will have some level of matching to off-target sites when running BLAST searches, the goal is to pick the primers that have the fewest off-target sites.

The protocol outlined below we will use BLAST to **compare** primers already screened and suggested for PCR by Consed or some other program (e.g. Primer3) and for which you have no other criterion (e.g. repeat location, product size, etc) to help you choose. It is not a method to select a primer pair *de novo* from a large genomic section. The goal of the BLAST search is to allow you to compare the alignments of the primers to various sites in the genome with the goal of choosing the “best” primer.

BLAST primer search protocol

Use the following set of guidelines when performing BLAST searches for off-target binding sites:

Search with the exact sequence of primer, using either FlyBase BLAST (<http://flybase.org/blast/>), or NCBI BLAST. In either case be sure to correctly set the specific species that matches your project and set the “Expect value” threshold to 10.

In addition:

For FlyBase BLAST in the Advanced Settings (optional) section:

Enter the following in the “Other options” text box to set the mismatch penalty: -q -1

For NCBI BLAST be sure to use blastn not megablast and set the following:

Be sure to restrict the search to the proper species’ whole genome assembly

Set Word size: 7

Turn **OFF** Automatically adjust parameters for short input sequences

Set match/mismatch score to: 4,-5

Turn **OFF** all filtering and masking

After running the BLAST primer search protocol follow these guidelines when comparing two different primers and the likelihood of off-target priming:

1. The last 2 bases at the 3' end must match for DNA priming to occur. Thus, primers with off-target alignments where the last 2 bases **do** align should be avoided if at all possible. You may run across situations where all primers under consideration have alignments of this type, in this case one would prefer the primer with the fewer number of such off-target alignments.
2. If multiple primers are still possible after elimination by criterion 1 then the length of the alignment to off-site targets: For primers where the last two bases do not match, it is still possible for these primers to bind to off-target sites with enough efficiency that if effects to concentration left to anneal to the correct site. Thus primers with long matches (say 80% of the length or more) to off-target sites should be avoided. Thus, when comparing two primers, choose the one with fewer of these long alignments even if the last two bases do not align.
3. If 1 and 2 still do not indicate the primer to order GC Content should be examined: primers with higher GC content can bind tighter than longer primers that are AT rich. Thus when comparing two primers with equally long alignments to off-site targets, the primer with the lower GC content would be preferred over the primer with the higher GC content.
4. If all three above criterion fail to pick a primer you may flip a coin or choose some other appropriately random method.