

## CALLING ADDITIONAL READS FOR HYBRID ASSEMBLIES

1. Select an appropriate primer pair for each reaction.

- a. Check primers for "uniqueness" using BLAST
- b. If possible keep the distance between the two primers to less than 1000 bp.
- c. Check that your chosen primer has a GC content of ~40-60%.
- d. Ensure that the estimated melting temp. of your chosen primer is >50°.
- e. Very difficult regions may require you to drop the GC content of the primer to less than 40%; in this situation you must keep the melting temperature above 50° by increasing the length of the primer.
- f. For each location, pick 2 different PCR primer pairs. Ideally each left primer will be compatible with both right hand primers (i.e. all 4 pairwise combinations will be listed as viable pairs by Consed).

2. Naming conventions for the reads.

The names of the reads will be constructed using various data bits to help you identify the properties of the read.

--- The reaction name should start with the prefix (**<GI code>**) as specified by the Genome Institute (e.g. XBAE-). This part can be safely ignored as it has information useful only to the Genome Institute.

-- The next prefix (**<faculty name>**) corresponds to the GEP faculty who claimed the project

The name of project (**<project name>**) follows the GI code (e.g. DBIA1495001)

Chemistry (**<chemistry>**) is indicated with an underscore:

--- Reads sequenced using 4:1 chemistry are identified with an attachment of "\_t"

--- Reads called using dGTP chemistry are identified with an attachment of "\_g"

--- Reads using normal BigDye are identified with only an underscore

Oligo for Sequencing

--- When using Consed, each oligo "accepted" is given a name based on the project name and given a number. Thus, the first oligo picked for project "DBIA1495001" would be called DBIA1495001.1, the next oligo would be DBIA1495001.2 etc. The oligo number corresponds to the digits after the period.

--- The number of the oligo used for sequencing is added to the reaction name after the chemistry (example file names with "\_7" [oligo 7 with BigDye chemistry] and "\_t7" [oligo 7 4:1 chemistry]).

Reaction names when using PCR products for templates (instead of subclones) are indicated by adding the letters "PCR" after the project name. The PCR reaction is described by the oligo numbers for the two oligos used for PCR (**<oligo #1>** and **<oligo #2>**), separated by either a "c" or "g" to indicate either a clone template or a genomic DNA template. For all GEP projects only "g" will be used (i.e. all PCR will use genomic DNA.)

The end of the read name will contain either b or g. This letter has no meaning in this context but remains for backward compatibility. The final number (**<reaction number>**) is used to indicate the number of times the reaction has been processed: reactions that have been processed the second time will end with .b2 or .g2 and reactions that have been processed the third time will end in .b3 or .g3.

The full reaction name can be described like this (not all parts will be present in all names):

**<GI code>**-**<faculty name>****<project name>**PCR**<oligo #1>****<c or g >****<oligo #2>****<chemistry>****<oligo number for sequencing>****<.b or .g>****<reaction number>**