



MYcroarray – info@mycroarray.com

Guidelines for submitting sequences for custom MYbaits kit design

December 2016

To submit your sequences for design, we accept sequences in either FASTA format or as coordinates from a reference genome (see below). Please note that we will design baits from ALL sequences or coordinates that you provide. If you only want specific regions of those sequences in the baitset (**e.g. exons only**), please first curate your targets to only include those specific regions of interest.

Option 1.) FASTA sequences

Baits of your desired length & tiling density will be designed along the full length of your provided sequences. Edit your sequence names according to the following guidelines:

- 1) *Allowed characters are letters, numbers, and dashes "-"*
(e.g. NO underscore "_", periods ".", commas ",", colons ":", pipes "|", brackets "[]" or "{ }", parentheses "(", slashes "/" "\", equal signs "=", or other punctuation "# ? !")
- 2) *No spaces (so please remove FASTA descriptions, or incorporate the necessary info into the name)*
- 3) *Length of sequence name 50 characters or less, if possible*

It is strongly recommended, though not required, that you incorporate the applicable species and locus into the sequence names, since baits will be named according to their parent sequence.

You may provide alignments if you wish, though all gaps will be ignored for bait design.

Singleton and/or short stretches of N's will be replaced with T's to facilitate bait design in these regions. Longer stretches (e.g. 10+ N's) will be skipped over during bait placement. Ambiguities (e.g. Y/M/R/S/W/K) are allowed, but will be replaced by a SINGLE random candidate base for manufacturing, since we can only synthesize A/T/C/G bases. If you wish to have ALL base options for that position represented in the baits ("mixed bases"), this must be done at the design stage, and will count as separate bait sequences.

Option 2.) Genomic coordinates

We can extract your sequences of interest from a reference genome using coordinates. Please provide us with a specific link to the exact reference genome that you are using, otherwise the coordinates will be incorrect. Please provide a text file in the following format (the locus direction, e.g. "+:" or "-:", is optional):

```
chr1:100345-134000:+  
chr5:13424324-14400821:+
```

Alternatively you can provide a tab-delimited file with these values in separate columns with clearly labelled headings. Please note that these coordinates are along the specific chromosome (i.e., the numbering restarts at the beginning of each chromosome). Please double-check that the names of your targets correspond to the chromosome/contig/scaffold names in the reference genome.

Sequences or coordinates can be submitted via email to your MYbaits representative.

Please compress large files, or post them on a file-sharing service (e.g. Dropbox or FTP).