

Submission Guidelines for Custom myBaits Design

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.

I. SEQUENCES

Acceptable: **FASTA DNA sequence format, in plain text file**

Names

- All sequence names must be fully unique
- Allowed characters are letters, numbers, and dashes "-" ONLY (no other characters should be used)
- Spaces and underscores will be replaced with dashes
- Name length 50 characters or less
- Recommended to incorporate species/locus names

Sequences

- Allowed characters are IUPAC bases
- Alignment gaps ("-") may be present, but will be ignored during bait design

Note regarding non-ATCG bases: 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 ONE random candidate base for manufacturing, since we only synthesize A/T/C/G bases.

II. GENOME COORDINATES

Acceptable: **BED ("Browser Extensible Data") format, in plain text file**

BED file format details available at: <https://genome.ucsc.edu/FAQ/FAQformat.html#format1>

- Provide link/copy of exact reference genome, otherwise coordinates will be incorrect
 - Names of chromosome/contig/scaffolds must match genome entry names exactly
 - Plain text file only (tab-delimited)
 - Do not submit spreadsheet/excel files
-

**Contact your myBaits representative
for instructions on how to submit your prepared files**