**Candidate Marker Submission for Targeted Genotyping**

There are 3 files required to prepare a set of candidate markers for submission.

1. *Guide\_For\_Marker\_Submission\_and\_Specification\_Form\_v0.9.1.docx* (this document)
2. *Format\_Instructions\_For\_Marker\_Submission\_With\_Examples\_v0.9.1.xlsx*
3. *Template\_File\_For\_Marker\_Submission\_v0.9.1.csv*

Please read the instructions provided in this document, and the example entries provided in *Format\_Instructions\_For\_Marker\_Submission\_With\_Examples\_v0.9.1.xlsx*.

Add your marker submissions to the provided csv file template, *Template\_File\_For\_Marker\_Submission\_v0.9.1.csv*.

**Please also fill out the Specification Form included at the end of this document**.

**Instructions**

The marker submission template csv file consists a header row defining the columns to be filled out for each candidate marker submitted.

There are 12 columns in the template, although not all columns are required in all cases.

Those columns which are optional are marked as optional in the column descriptions.

Column 1: **MarkerName**

Name used for identifying and reporting markers.

Each candidate marker must have an entry in this column. All marker names must be unique within the marker set.

Column 2: **TargetSequence**

Defined allele sequences with flanking sequence on both sides.

See the document *Format\_Instructions\_For\_Marker\_Submission\_With\_Examples\_v0.9.xlsx* for examples of the required format.

DNA strand orientation must be 5' to 3', with allele variants and flanking sequence defined on the same strand.

Flanking sequence consists of DNA nucleotides bases, including IUPAC ambiguity codes, in upper case, with no spaces or line breaks allowed.

"Reference allele" and "Alternative allele" are defined at the appropriate position in the TargetSequence in square brackets [REF/ALT], as per the example [C/A].

Non ACGT IUPAC symbols cannot be present within the square brackets except for '-' when used for deletion variant of indel markers.

Indel type markers may be represented as a standard insertion and deletion pair such as [T/-] or [-/TTA], or may be represented by a pair of distinct sequences such as [GA/AAT] or [GGAC/TGGAT].

Indel type allele definitions must not contain any common prefix or suffix bases, for example [TAGTGA/TAGA] is not permitted, as it contains common prefix and suffix.

Only a single instance of '-' can be present within the square brackets and cannot be combined with other characters to make an allele sequence.

If possible, provide 150bp nucleotide sequences from right and left ends of the Variant - ATCGAACGTT….150bp [REF/ALT]ACTTGCA…... 150bp

Provide a minimum of 100bp total sequence length to allow a proper probe design

**Note**: Allow only a single variant specified within the sequence - i.e. only a single pair of square brackets [REF/ALT], and only 2 alleles within the brackets.

Definitions with more than 2 alleles [REF/ALT/ALT] cannot be processed. These can be represented as separate markers with one [REF/ALT] combination per marker submission row.

Column 3: **ReferenceGenome**

Optional. When available, provide the name, including version information, of the Reference genome used to provide position information for the marker.

For the Reference genomes(s) included here, provide a download link in the accompanying specifications section at the end of this document.

Column 4: **Chrom**

Optional. When available, provide the chromosome/contig from the Reference genome listed in Column 3.

Column 5: **ChromPosPhysical**

Optional. When available, provide the position (bp) on the chromosome/contig listed in Column 4.

Column 6: **ChromPosGenetic**

Optional. When available, provide the genetic position (cM) on the chromosome listed in Column 4.

Column 7: **VariantAllelesDef**

Specify here the Ref and Alt alleles of the marker in the same format used in the TargetSequence in Column 2, [REF/ALT].

Check that the defined Ref and Alt alleles correspond exactly the definition included in Column 2, and that the defined alleles are on the same strand as the flanking sequence.

Column 8: **MarkerType**

Use one of the two descriptions, SNP, or INDEL.

Column 9: **EssentialMarker**

Optional. Indicate those markers which are considered to be essential, and should be included in the selected marker panel regardless of design quality or other selection and optimisation criteria.

Markers which are considered essential should be marked with YES in this column, otherwise NO.

Column 10: **MinorAlleleFrequency**

Optional. If applicable, a MAF value can be provided for use in marker selection.

If MAF values are provided, they must be given for all candidate markers with the exception of those marked as essential in Column 9.

Column 11: **Quality**

Optional. If applicable, include a quality score for use in marker selection.

The quality score should consist of values between 1.0 and 10.0, with 1.0 indicating the best quality.

If quality scores values are provided, they must be given for all candidate markers with the exception of those marked as essential in Column 9.

Column 12: **Comments**

Optional. Additional comments may be included here.

**DArTseq Markers**

DArTseq marker can be used as input for all or part of the candidate marker set.

When submitting DArTseq markers, please provide the list of markers in a separate .csv or .xlsx file. It is important to also include the ‘2-row’ format of the DArTseq SNP marker report which accompanied the report file from which you selected your candidate DArTseq markers. The ‘2-row’ format report file is required to perform correct marker design.

The ‘2-row’ marker file can be recognised by the naming convention of alternate Ref/Alt marker rows in the SNP data table:

|  |
| --- |
| AlleleID |
| 25317125|F|0--27:G>T |
| 25317125|F|0-27:G>T-27:G>T |
| 25332532|F|0--26:A>T |
| 25332532|F|0-26:A>T-26:A>T |

**Specifications**

Please provide here the general specifications for your required marker set.

The marker design process produces various quality parameters based on flanking sequence composition, and target region copy number (if a Reference genome is available).

These quality parameters will be used, in combination with optional values provided in the EssentialMarker, MinorAlleleFrequency, and Quality columns to choose a set of markers to make up the panel.

When an optimal selection is required, providing an excess of suitable candidate markers will improve the outcome of the marker selection process.

Optionally, if requested, marker design and design quality determination can be performed on the full set of candidate markers, and then returned to you to perform your own final selection of the optimal marker subset for panel creation.

Please specify below how you would like the marker selection to be performed. For example, when marker position information is available, markers can be selected to provide an optimal spread across the genome, in addition to meeting specific quality criteria.

|  |
| --- |
|  |

**Marker Panel Details**

|  |  |
| --- | --- |
| Total number of Markers Required: |  |
| Number of candidate markers provided in submission file: |  |
| Number of DArTseq candidate markers provided: |  |

**Reference Genome download links**

(for any reference genome listed in the submission file, provide a link here)

|  |  |
| --- | --- |
| Reference Genome Name | Download Link |
|  |  |
|  |  |
|  |  |