AgriSeq targeted GBS custom panel design for agrigenomics applications

Cost-effective customization made fast and easy

AgriSeq targeted GBS solutions

The Applied Biosystems™ AgriSeq™ targeted genotyping-by-sequencing (GBS) solution allows high-throughput analysis of plant and animal genotypes in a flexible and cost-effective manner. Leveraging highly efficient multiplex chemistry, hundreds to thousands of genetic loci can be simultaneously targeted and efficiently amplified in a single reaction.

Customized panels designed to target your relevant markers

AgriSeq chemistry is designed for analysis of SNP panels containing 50–5,000 markers. We have a dedicated team of agrigenomics bioinformatics scientists who have experience designing a range of successful panels for various organisms. Our targeted GBS solution for plant and animal genotyping offers the ability to customize your genotyping content.

Overview of the design process

The AgriSeq targeted GBS panel design process allows you to create an entirely new design or modify an existing AgriSeq panel.

- A structured, collaborative process between you and our agrigenomics bioinformatics team (Figure 1)

- Easy to iterate design by adding, replacing, or removing markers as needs change

- Compatible with all types of markers, including SNPs, MNPs, indels, translocations, and inversions

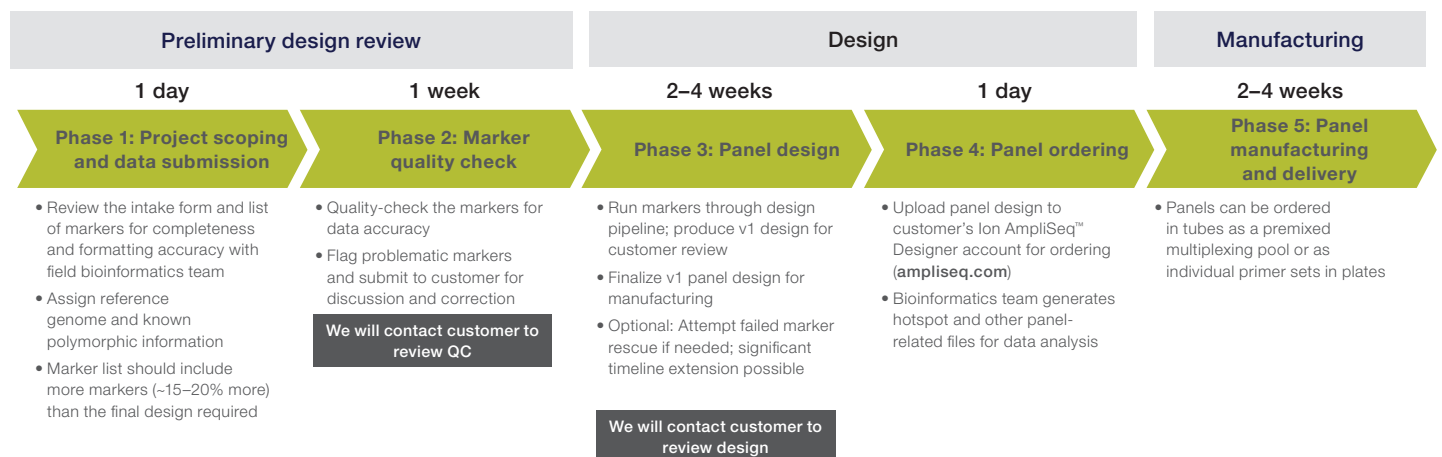- Streamlined process from design to manufacturing can be completed in 4–8 weeks

Preliminary design review		Design		Manufacturing
1 day	1 week	2–4 weeks	1 day	2–4 weeks
Phase 1: Project scoping and data submission	**Phase 2: Marker quality check**	**Phase 3: Panel design**	**Phase 4: Panel ordering**	**Phase 5: Panel manufacturing and delivery**
• Review the intake form and list of markers for completeness and formatting accuracy with field bioinformatics team • Assign reference genome and known polymorphic information • Marker list should include more markers (~15–20% more) than the final design required	• Quality-check the markers for data accuracy • Flag problematic markers and submit to customer for discussion and correction **We will contact customer to review QC**	• Run markers through design pipeline; produce v1 design for customer review • Finalize v1 panel design for manufacturing • Optional: Attempt failed marker rescue if needed; significant timeline extension possible **We will contact customer to review design**	• Upload panel design to customer's Ion AmpliSeq™ Designer account for ordering (**ampliseq.com**) • Bioinformatics team generates hotspot and other panel-related files for data analysis	• Panels can be ordered in tubes as a premixed multiplexing pool or as individual primer sets in plates

Figure 1. AgriSeq targeted GBS panel design workflow.

Thermo Fisher
S C I E N T I F I C

Optimize your panel for your genotyping application

Many genotyping applications (e.g., molecular breeding and parentage verification) require the ability to interrogate a variety of marker types, including SNPs, MNPs, indels, inversions, and copy number variations (transcript-derived targets, organelle targets, and STRs are not currently compatible with the AgriSeq design). The AgriSeq targeted GBS technology allows you to include all desired markers in a single panel; however, the target locus information required to design these markers differs.

- All required information for design is collected in the 10-column "Marker Submission Form"

- SNPs, MNPs, and small indels require completion of the marker submission form (Figure 2A)

- Work with our field bioinformatics team to complete the marker submission form if assistance is needed

For large structural rearrangements (e.g., large indels, translocations, inversions, etc.) a slightly different submission format is needed (compare Figures 2B and 3).

- Large indel submission is similar to small indel submission except that allele sequences cannot be copied and pasted into a spreadsheet cell due to size

- An additional file must be provided with sequences

The AgriSeq Design and Pilot Workbook has 4 tabs:

- **Tab 1:** Read me—this tab is informational and explains the other 3 tabs in the workbook

- **Tab 2:** Intake form

- **Tab 3:** Marker submission form

- **Tab 4:** Sample submission form (for wet-lab pilots only)

Tab 2: Intake form

This tab captures the customer information and project information to help ensure alignment of project goals and deliverables. Some answer fields will change colors to warn that the information provided should be further discussed with our team before proceeding. The information requested is described below.

- **Customer information**—identifies the primary point of contact for the project. The user must also sign up for an ampliseq.com account to order panels.

- **General information**—captures the high-level information of the experimental design: description of the experimental objective, the organism, the ploidy, and the desired number of markers in the final panel. We request that customers supply 10–20% more markers than desired for their final panel to allow for drop-off in the design and wet-lab phases.

A

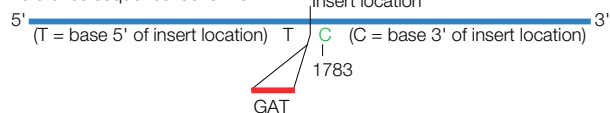| Marker ID	Chromosome	Start	End	Reference	Variant	Strand	Type	Priority	Sequence
Marker-1	CHR01	87964	87964	A	C	Plus	SNP	1	AAAACCAAAAAGCACACTGGAC[A/C]CCCCACGCAGCTGGTTTGACAT
Marker-2	CHR28	244703	244704	TG	AA	NA	MNP	2	CAAAGATACTGTAAAGTTAATTATTATC[TG/AA]GGCAGCAAGGCCA

B

Deletion
Reference sequence looks like:

Start = 36 End = 39
5' ————————————— 3'
ACTA
36   39

Insertion
Reference sequence looks like:
Insert location
5' ————————————— 3'
(T = base 5' of insert location) T   C   (C = base 3' of insert location)
1783
GAT

Marker ID	Chromosome	Start	End	Reference	Variant	Strand	Type	Priority	Sequence
Marker-3	Chr_01	36	39	ACTA	–	Plus	DEL	2	GCT[ACTA/-]CGA
Marker-4	Chr_03	1783	1783	–	GAT	Plus	INS	1	AAG[-/GAT]CTG
Marker-5	Chr_11	721279	721279	T	–	Minus	DEL	1	TTT[T/-]GAC

Figure 2. Marker submission form for SNPs, MNPs, and small indels. (A) Example of a completed marker submission form for SNPs, MNPs, and small indels. The text in red indicates the hotspot. The headings of the form are described on the next page. **(B)** Example of completed Marker Submission Form for small indels. Red lines indicate the sequence to be inserted or deleted. The text in red indicates the hotspot. For insertions, give the 3´ insertion location as both the "Start" and "End" sequence (shown here as "C" in green at position 1783).

A. Marker Submission Form								
Marker ID	**Chromosome**	**Start**	**End**	**Reference**	**Variant**	**Strand**	**Type**	**Priority**
Marker-6	Chr_01	36	41	NA	–	Plus	BIGDEL	2
Marker-7	Chr_03	1783	1783	–	NA	Plus	BIGINS	1
B. Longindel.txt								
Use brackets to contain alleles and provide 200 bp flanking: Marker-6 GCCAACAGTA[ACTAGC …/-]CATAGCTCTG Marker-7 TCGATCCGAT[-/GATGAT …]GAGCGATCGA								

Figure 3. Example of completed marker submission form for large indels. For large indels, in addition to the marker submission form **(A)**, please also provide sequences in a separate Longindel.txt tab-delimited file **(B)**. Names for these marker IDs must match their spreadsheet counterparts. Note: Ellipses (…) are used here for brevity's sake, but we require that the full sequence be provided.

- **Performance requirements**—identifies the customer's expectation of the panel's performance and any must-have markers. While we routinely achieve marker call rates in the mid–high 90% range, we cannot guarantee design success of every marker.

- **Bioinformatics information**—collects the genomic information and marker types for the desired panel. We prefer to use the latest build of the organism's reference genome for completeness, but we can also design solely from the supplied contig sequences. In addition, incorporating known SNP information from dbSNP or other user-supplied SNP data can help to improve the primer design by avoiding polymorphic regions.

- **Analysis considerations**—this sections is only relevant for wet-lab pilots. Customers can indicate their interest in identifying not only hotspot variants but any other novel variants as well. In addition, the user can request the types of files they want (FASTQ, BAM, VCF, and genotype tables), and their preferred method of data delivery. VCF data and genotype calls are based on optimized Torrent Variant Caller (TVC) software—use of open source software is not recommended for analysis.

Tab 3: Marker submission form

The information requested on this tab enables our GBS bioinformatics design team to design the GBS panel for your species of interest. All fields marked with an asterisk (*) are mandatory (Figures 2 and 3). Other fields can be filled with "NA".

- **Marker ID***—identifies the marker. Names should be shorter than 20 characters long, and we prefer alphanumeric values (names cannot contain spaces or special characters).

- **Chromosome***—each submitted marker should have a unique chromosome location and be consistent with the reference genome.

- **Start location***—captures the reference allele position in the reference genome for all marker types.

  – For SNPs, MNPs, and deletions—position on the genome

  – For insertions—use the coordinate position of the base immediately 3′ to the insertion

- **End location***—indicates the variant allele position in the reference genome for all marker types.

  – For SNPs—same as the "Start" location (Start = End)

  – For MNPs and deletions—end position in the reference genome

  – For insertions—use the coordinate position of the base immediately 3′ to the insertion; same as the "Start" location (Start = End)

- **Reference allele***—indicate the allele to be called the reference allele; it should match the reference allele in the context sequence or the reference genome. No degenerate bases allowed.

- **Variant allele***—indicate the allele to be called the variant allele. It should match the variant allele indicated in the context sequence. No degenerate bases allowed. By default we treat each marker as biallelic, but for triallelic (multiallelic) markers, indicate each allele as a separate marker. This helps to get the right genotype call but has no impact on the design.

- **Strand**—values are "Plus", "Minus", or "NA". Marker positions, alleles, and context sequence should match the correct strand on the reference genome. This is crucial for marker validation as well as getting the correct genotype call. Regardless of the strand indicated, TVC provides variant calls based on the forward strand (regardless of how the context sequence maps to a genome). If you are not sure, enter "NA".

- **Marker type**—captures the type of marker. This information helps us to follow the right marker validation, design, and hotspot file generation strategies. SNP = single nucleotide polymorphism; MNP = multi-nucleotide polymorphism; DEL = deletion; INS = insertion; BIGDEL = deletion longer than 50 nucleotides; BIGINS = insertion longer than 50 nucleotides.

- **Priority**—highlights the marker importance. The bioinformatics team will attempt to design all markers but will pay special attention to priority 1 markers. Please be aware that not all markers are guaranteed to yield a design, and additional rescue efforts can cause delays.

  1 = must-have

  2 = desirable

  3 = not a priority

- **Sequence***—the context sequence is described as a 401-mer nucleotide sequence from the plus strand centered on the target SNP. The alleles of the marker are indicated in brackets [ ] at the 201st position.

An example of a 401-mer sequence for a SNP submission:

CGTGACAGCATCAGCATAGATAGAAGATAGAATGCTGACG
TGACAGCATCAGCATAGATAGAAGATAGAATGCTGACGTG
ACAGCATCAGCATAGATAGAAGATAGAATGCTGACGTGAC
AGCATCAGCATAGATAGAAGATAGAATGCTGACGTGACAG
CATCAGCATAGATAGAAGATAGAATGCTGACGTGACAGCA
[T/C]CGTGACAGCATCAGCATAGATAGAAGATAGAATGCTG
ACGTGACAGCATCAGCATAGATAGAAGATAGAATGCTGAC
GTGACAGCATCAGCATAGATAGAAGATAGAATGCTGACGT
GACAGCATCAGCATAGATAGAAGATAGAATGCTGACGTGA
CAGCATCAGCATAGATAGAAGATAGAATGCTGACGTGACA
GCA

If the context sequence originated from the minus strand, this should be indicated correctly in the "Strand" column. For shorter context sequences, please contact your sales representative to get more information. For long indels, please provide sequences in a separate tab-delimited text file (Figure 3). Marker names in the text file should match the marker names in the marker submission form.

Tab 4: Sample submission form (for wet-lab pilots only)

This tab captures sample information for projects with a wet-lab component. We ask for the number of unique samples, the identity of any replicates or samples with known relationships (parents/offspring), and the method of sample preparation and quantitation, as well as desired disposition of the samples after the pilot project. Lastly, while it is not imperative to submit samples with known genotypes, having orthogonal or truth data can help optimize our genotype calling software to improve marker call rates for your particular samples and AgriSeq panel.

applied biosystems

Sources of genomic information for AgriSeq targeted GBS panel design

Variants for AgriSeq custom panels may be selected from several sources, including but not limited to the following:

• dbSNP

• Markers identified by sequencing or GWAS studies

• The Applied Biosystems™ Axiom™ Genomic Database or other array databases (Figure 4)

• Markers personally chosen by you

• Other array content description files

Information related to marker type and location within the reference genome can often be found in array databases online (e.g., Axiom Array database; URL). An example of how to transfer the data from existing technologies to the AgriSeq Marker Submission Form is shown in Figure 4.

	1	2	3	4	5	6	7	8	9	10
	Marker ID	Chromosome	Begin location	End location	Reference allele	Variant allele	Strand	Marker type	Priority	Sequence
AgriSeq marker submission form	Target-1	CHR01	87964	87964	A	C	Plus	SNP	1	…GTCCCACG CAGCTGGTTT GACATCCCCA CGCAGCTGG TTTGACATCC GA[A/C]CCCC ACGCAGCTG GTT…
Corresponding information from Axiom Annotation File	Customer to provide	Column 4 "Chromosome"	Column 5 "Physical Position"	Column 5 "Physical Position"	Column 8 "Allele A"	Column 9 "Allele B"	n/a	n/a	Customer to provide	Column 7 "Flank"
Corresponding information from Illumina Manifest File	Customer to provide	Column 10 "Chr"	Column 11 "Map Info"	Column 11 "Map Info"	Column 4 "SNP"	Column 4 "SNP"	n/a	n/a	Customer to provide	Column 17 "Source Seq"

Figure 4. Example of microarray description file from an array database.

Find out more at **thermofisher.com/agriseq**

Thermo Fisher SCIENTIFIC