Guidance notes

SeqSNP service guidance notes

For Research Use Only. Not for use in diagnostic procedures.

BIOSEARCH™ TECHNOLOGIES

GENOMIC ANALYSIS BY LGC

Guidance notes

SeqSNP service guidance notes

We are pleased to provide the following guidance on how LGC, Biosearch Technologies will work with you regarding your SeqSNP service project.

The SeqSNP service includes:

- Dedicated project manager
- LGC plant sample collection kit(s) (optional)
- DNA extraction service (optional)
- SeqSNP assay design and SeqSNP kit production
- Library preparation and Illumina sequencing
- Variant calling

The following topics are covered in this document:

SeqSNP is a targeted genotyping by sequencing (tGBS) service by Biosearch Technologies, which allows for genotyping of SNPs and small insertions/deletions (InDels) via a midplex genotyping platform. It can provide an alternative for screening via fixed arrays, due to increased flexibility in marker selection and ease of sample number scalability.

The SeqSNP service is all-inclusive, allowing for DNA extraction (alternatively, customers can submit their extracted DNA), SeqSNP assay design (*in silico* design of oligonucleotide probes flanking the SNPs of interest), SeqSNP kit synthesis (generation of the kits to allow for the interrogation of the SNPs), and Illumina sequencing and bioinformatics to allow for the calling of alleles.

Please enquire for a free-of-charge consultation with our sequencing specialists at seqsnp@lgcgroup.com.

Guidance notes

SeqSNP service guidance notes

Section 1: Project workflow

Preliminary work: SeqSNP assay design and SeqSNP kit production

a. Customer assigned a dedicated project manager
b. Verification of customer-provided SNP information
c. SeqSNP assay design
d. SeqSNP assay design verification
e. Customer confirmation of assay design, i.e. final selection of SNPs
f. Synthesis of SeqSNP kit (the oligo probe library)

Phase 1 – Sample preparation

g. Receipt of samples/tissue material
h. DNA extraction (if required)
i. Quality control of submitted or extracted DNA

Phase 2 – Sample processing

j. Library preparation including indexing and quality control
k. Illumina sequencing

Phase 3 – Data analysis

l. Processing of raw reads and quality trimming
m. Sequence alignment and variant calling
n. Data supplied to the customer on USB stick or hard disk drive. Analysed results will consist of a table (VCF and Excel format) containing SNP genotypes. Raw sequence data and alignments are also provided.

Section 2: Project submission requirements

Part A: Documentation and data requirements

Biosearch Technologies requires the following documents and data in order to commence work on your project:

- Purchase order or signed quotation.
- A completed NGS sample submission form. Please provide information regarding the DNA extraction method used and the concentration (ng/µL) of your submitted samples. Also, please provide information on the ploidy of the species to be studied, and any details as to the stage of the breeding programme the samples are taken from, if applicable.
- A file containing **SNP information** (various acceptable formats).
- A file containing the **reference genome sequence**, or a link to the reference genome.
- It is important that the input data (i.e. the SNP file and reference genome) is checked and is correct before submission. Unfortunately, Biosearch Technologies cannot accept responsibility for any submitted sequence data which contains incorrect information.

- If you are a new customer, please also register with our sequencing online shop to create your own uniqueBiosearch Technologies customer number. We cannot begin processing of your samples until a customer number is available.

Please send all documents to seqsnp@lgcgroup.com **before** shipping your biological samples.

Please be aware that if you do not supply the above documents, we cannot process your project, and the delivery of results will be delayed.

Part B: Sample requirements

The minimum number of samples that can be submitted for a SeqSNP service project is 48; however we recommend a minimum of 96 samples, due to the scales of the kits. Samples for a SeqSNP project can be submitted in the format of extracted DNA (see Note (i)) or tissue samples (see Note (ii) and Note (iii)). DNA extraction is included in the project cost for specific sample types as detailed in Note (ii) and Note (iii). Both DNA samples and tissue samples should be submitted in a 96-well plate format. Samples submitted in tubes may incur a surcharge for sample transfer (excludes blood or buffy coat samples).

Please make sure to contact the SeqSNP team before sending any samples to allow for any recommendations regarding shipment.

If you are working with non-standard sample types, species or delivery formats that are not covered in this document, please note that additional costs may apply.

Note (i): Submitting DNA samples

When submitting DNA for a SeqSNP project, samples should be submitted in 96-well plates. There is no requirement to leave any wells empty as controls, and samples can be formatted in rows or columns. Each plate submitted should be labelled, and the corresponding sample information detailed in the sample submission form (see Documentation requirements section).

We require a minimum of 20 µL of DNA per individual sample, at a minimum concentration of 30 ng/µL (quantification by Qubit/PicoGreen).* DNA must be submitted in solution. We require high molecular weight DNA (average fragment length >10 kb), RNA-free and eluted in Tris/TE (5mM, pH: 8.5). Samples that do not meet the quality requirements may be refused. DNA can be shipped on ice or at room temperature.

Biosearch Technologies recommends performing DNA extraction using our sbeadex™ chemistry, that delivers nucleic acids of high yield, purity and quality.

*If your DNA samples are less concentrated, please contact Biosearch Technologies for further details. It may be possible for us to obtain good quality data with slightly less concentrated samples.

Note (ii): Submitting plant tissue

To submit plant leaf tissue, first determine whether your plant species has been previously extracted by Biosearch Technologies by consulting our current plant species list.

a. **If your plant species is listed**:
   Please submit plant tissue using Biosearch Technologies Plant Sample Collection Kit, following the sampling instructions provided with the kit. For SeqSNP projects, we require you to submit 6 to 8 leaf discs per sample. The number of leaf discs required to yield sufficient DNA is highly dependent on the age of the leaf sample and hence how rapidly the cells are dividing.
   Please note: As DNA extraction efficiency is highly crop and sample-type dependent, it is possible that our extraction protocol may require optimisation to ensure sufficient DNA is obtained. This may involve a short pilot study of your species. If your plant samples are precious, and further sampling at a later date is not possible, please discuss this with us prior to commencement of your project. It may be advisable for us to optimise the extraction protocol on a small set of test samples, prior to proceeding with the full set. Please note that all 96-well plates can be used for plant tissue, and there is no need for empty wells. Each plate submitted should be labelled, and the corresponding sample information detailed in the sample submission form (see Documentation requirements section).

b. **If your plant species is not listed,** you can:
   i. Submit extracted DNA for your SeqSNP project (see Note (i) for further details).
   ii. Discuss a pilot extraction project with specialists at Biosearch Technologies. Protocol development is not included in the SeqSNP service project price. Please note that standard turnaround times and conditions may not be achievable if bespoke protocols are required for your project. If you are unsure, please discuss requirements for your species with us prior to commencement of your project.

Guidelines for sampling plant tissue

- Please ensure sampled tissue is of a consistent age across all plants sampled. For older plants with mature leaves, a larger number of leaf punches may be required (e.g. 7-9) than if young leaf tissue is sampled (e.g. 4-6).
- Please ensure that sampling is performed on consistent positions of each plant.
- Avoid midrib and axillary veins on leaves.
- Sample leaf material at L3/L4 stage (i.e. 2-3 leaves on the main shoot, and third/fourth leaf has appeared).

Guidelines if using Biosearch Technologies Plant Sample Collection Kits

- Always re-package the plate in the corresponding box; the box and plate have corresponding barcodes.
- Use a fresh leaf punch and mat for each collection kit.
- Avoid puncturing the plastic bag when inserting the completed kit into the bag.
- Do not re-open the plastic bag containing the completed sample kit.

Once the collection has been completed, store all collection kits in the dark at room temperature.

Note (iii): Submitting non-plant tissue

If you would like to submit non-plant tissue samples such as whole blood, buffy coat or solid tissue for extraction, **please discuss this with Biosearch Technologies prior to the commencement of your project. Your project manager will advise you of sample requirements specific to your project.**

Part C: Shipping conditions

All samples for an individual SeqSNP project should be sent to Biosearch Technologies in a single shipment. It is essential to include your purchase order number or a signed quotation in your shipment to facilitate identification of the samples and to confirm your order for the service.

DNA samples can be shipped on ice or at room temperature. If tissue samples are being submitted for extraction, **please confirm the appropriate shipping conditions with us prior to shipping**.
If the Plant Sample Collection Kit is being used, please follow the detailed instructions provided in the kit manual.

Part D: Data submission requirements

For each SeqSNP project, a list of SNPs and information on the corresponding reference genome is required. Optionally you may provide a list of flanking SNPs. Please refer to Note (iv) and Note (v) for details.

Note (iv): SNP definition file

The SNP definition file defines the positions and identifiers of SNPs to be included in your SeqSNP service. The minimum number of SNPs is 48, however we recommend a minimum of 96 SNPs due to the scale of the kits.

We accept the following three formats:
1. VCF - Variant Calling Format, >= v4.0; (please see VCF file guidance)
2. Excel spreadsheet according to our template, with the following columns:
    a. SNP identifier
    b. Contig
    c. Position (1-based)
    d. Reference allele
3. BED – Browser Extensible Data, (please see BED file guidance) with the following columns:
    a. Contig
    b. Start position (0-based)
    c. End position
    d. SNP identifier

You can download examples for these file types here:
VCF file
Excel file
BED file

Please note that we cannot verify whether SNPs provided to us using the BED format are matching the reference genome sequence file. We therefore recommend submitting your SNP data in the form of VCF or Excel files (including the reference allele), which we will use to verify that the SNP definition file and the reference genome sequence file are matching. If given SNP positions are incorrect in this file, correct genotyping will not be possible.

We recommend that you submit 20-50% more SNPs than are to be included in your SeqSNP project. This will facilitate the SeqSNP assay design process. Before producing the SeqSNP kit, we will ask you to confirm your SeqSNP assay design with your final selection of SNPs.

In addition to the SNP definition file, you can (optionally) provide a file containing flanking, non-target SNPs. This file will be used in the SeqSNP assay design process to avoid overlaps with the oligonucleotide probe design, which would otherwise compromise target enrichment. Please use the same file format as used for the SNP definition file.

Note (v): Reference genome sequence file

Please provide a link to a publically available reference genome sequence, or supply a file with the reference genome sequence in FASTA format.

Example: Human genome assembly

If a reference genome is not available, shorter reference sequences can be used. However, we require at least 150 nucleotides surrounding the target SNPs (75 nucleotides on either flank). For special types of sequences (e.g. a transcriptome reference) please discuss this with your Biosearch Technologies Project Manager.

Section 3: Turnaround times

Standard turnaround times

- The standard turnaround time for a SeqSNP project is 2 weeks for projects with existing oligo libraries (excluding assay design) and approximately 8 weeks for new oligo libraries (including assay design):
    - SeqSNP assay design takes 1 week.
    - Producing the SeqSNP kit typically takes 4 to 6 weeks.
      It is therefore advisable to provide SNP and reference genome information as early as possible.
    - Biological/DNA samples can be sent in parallel with sending the SNP and reference genome information, or upon initiating the SeqSNP assay and kit production process.
    - Once the samples are received (and if required, the DNA extracted) and the SeqSNP kit is available, it takes approximately 2 weeks (see below for projects over 768 samples) to provide analysed data.

Guidance notes

SeqSNP service guidance notes

Non-standard turnaround times

- Please note that wet-laboratory turnaround times for projects over 768 samples will be longer than 2 weeks. Please enquire with your Project Manager for accurate turnaround times for your specific project.
- Longer turnaround times also apply for the SeqSNP assay design should there be special requirements, e.g. high complexity of the reference genome. Please be aware that samples that are shipped from outside Europe may have slightly extended turnaround times due to international transfer limitations.

Bespoke turnaround times

- Bespoke turnaround times are available on request. Please contact Biosearch Technologies to discuss this prior to commencing your project.

Section 4: Important points to note

1. It is the responsibility of the customer to provide accurate SNP and genome reference information at the start of a SeqSNP project.
2. If you wish to analyse the same set of SNPs for future projects please inform your Project Manager prior to submitting your SNP sequences. If possible please provide an estimation of how many samples; this may help in saving costs for future projects.
3. If submitting DNA, an additional charge may be incurred if DNA does not meet quality requirements. A subset of DNA samples will be quantified (using Qubit) by Biosearch Technologies, allowing us to determine if the samples are of sufficient concentration to be accepted.
4. It is possible that submitted DNA, or DNA extracted by Biosearch Technologies from submitted tissue, may meet the quality requirements, but that inhibitors may affect the quality of data obtained. Customers are advised that the presence of inhibitors is sample dependent and is outside of our control.
5. Pricing and conditions may vary depending on the technical requirements for specific organisms. Please enquire for a free-of-charge consultation with our sequencing specialists at seqsnp@lgcgroup.com.
6. Pricing is per sample and applies to a single work order (samples received in one batch). A minimum sample number of 48 samples applies.
7. Typical applications of data generated include genomic selection and genomic prediction. Please note that guidance on breeding programmes is not included in this service. Biosearch Technologies may be able to advise on particular publications that can provide guidance; please discuss with our Technical Support team at biosearch.techsupport@lgcgroup.com.
8. Customers will receive a table with the results of the SNPs identified according to those submitted in the SNP definition file. Further customer analysis of the raw FASTQ files may identify previously unidentified (*de novo*) SNPs, however, this additional bioinformatics is not included in the SeqSNP service.
9. Each sample will carry an individual index which allows sorting of the reads after sequencing.

10. Any DNA samples remaining following completion of the project will be stored, unless otherwise specified, for 1 year after the end of the project.
11. All data generated is stored, unless otherwise specified, for up to 1 year after data delivery.
12. If additional services are required (e.g. repeated genotyping of certain samples, repeated data delivery or retrieval, extension of storage time), additional costs may apply.

Section 5: Useful links and contacts

Plant sample collection kit	Please click this link for more details about Biosearch Technologies plant sample collection kit.
Plant species list	Please click this link to access a list of plant species that Biosearch Technologies has previously extracted DNA from.
DNA extraction services	Please click this link for more details on Biosearch Technologies DNA extraction services.
DNA extraction application form	Please complete this application form in advance of submitting your samples to Biosearch Technologies, if DNA extraction is required as part of your project. Completed forms should be sent to your Project Manager.
NGS sample submission form	Please complete this form in advance of submitting your samples to Biosearch Technologies. This form should be used to detail the well locations and DNA concentration information when submitting extracted DNA. It should also be used to detail the sample names of tissue that is being submitted for extraction (use the plate layout worksheets).
Sequencing online shop	If you are a new customer, please register online for your unique Biosearch Technologies customer account number.

Useful contacts

General enquiries:
seqsnp@lgcgroup.com

Price enquiries:
seqsnp@lgcgroup.com

Technical enquiries:
biosearch.techsupport@lgcgroup.com

Queries relating directly to your ongoing project should be sent to your assigned Project Manager.

The quality management system of LGC Genomics GmbH in Berlin, Germany is certified by DIN EN ISO 9001:2015.
Delivery of result: Scored genotyping data. Raw data can be provided on request.
Payment conditions: 30 days after date of invoice net.
Quotations are valid for 90 days, unless specified otherwise.

Offered net prices do not include shipping costs and taxes, where applicable. Customers within the European Union have to account for the VAT and need to provide the VAT identification number to LGC Genomics GmbH for correct invoicing.

All services offered are subject to LGC Genomics GmbH's Terms and Conditions.

Integrated tools.
Accelerated science.

@LGCBiosearch | biosearchtech.com

BIOSEARCH™ TECHNOLOGIES

GENOMIC ANALYSIS BY LGC