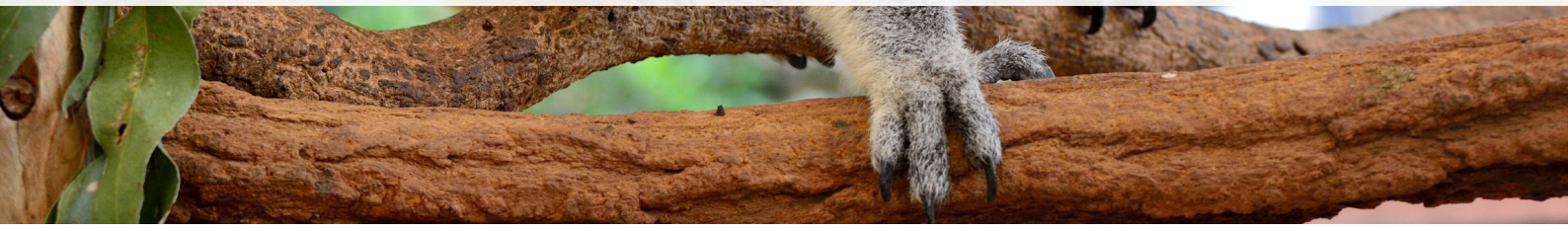
SERVICE GUIDE

Genotyping By Sequencing Service



CustomerCare@agrf.org.au
1300 247 301
www.agrf.org.au

SVG2305GBS_Service Guide_GBS

Service Guide:
GBS Service

Service Guide:
GBS Service

1.0 Overview

Genotyping-by-Sequencing (GBS) is a next generation library preparation and sequencing method that can be used simultaneously for both SNP discovery and genotyping. Using restriction enzymes to reduce genome complexity, GBS offers high-throughput, highly multiplexed sequencing. This approach is applicable to a range of biological questions such as population studies, linkage and QTL mapping and GWAS, and is suited to a wide range of organisms.

With AGRF's GBS method, samples undergo a double-digestion, followed by ligation with 48 unique barcoded adapters – this allows us to then pool 48 samples into a single library. Libraries then undergo a size selection and an indexing PCR. We use ten sequencing indexes, combined with the 48 barcoded adapters, and this gives us the capacity to have up to 955 samples uniquely indexed.

GBS is suited to any species and can be particularly informative in species where published genomic data is unavailable.

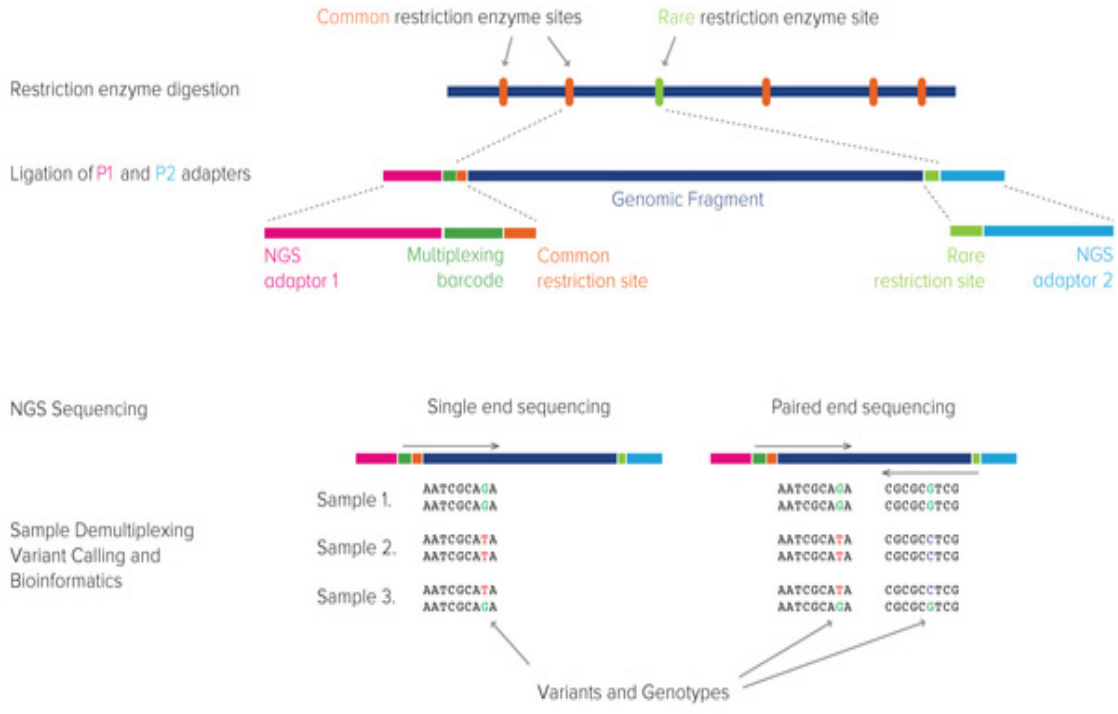


Figure 1: GBS workflow.

2.0 Submission types for the GBS Service

Our GBS workflow is usually carried out in two stages:

Stage 1: GBS Establishment Service

To enable efficient library preparation, we offer establishment services for any new project prior to batch processing.

Standard Establishment

This protocol determines a suitable enzyme combination for library preparation. Three samples are pooled, digested with eight different restriction enzyme combinations, and libraries are made and examined by electrophoresis for the combination least likely to yield repetitive sequences. We use PstI or EcoRI in combination with MspI, MseI, NlaIII and HpyCH4IV.

Stage: 2 Batch Processing

Following the establishment of the optimal library preparation method, GBS project samples are processed in batches of 47 (plus a negative control), 95 (plus a negative control), 143 (plus a negative control) or 191 (plus a negative control) per submission which enables very high sequencing economies. Sequencing data is analysed through our bioinformatics pipeline.

AGRF can perform DNA extraction if you require, please discuss this with your Account Manager.

2.1 Technical Considerations

- The standard establishment provides an indication of the optimal enzyme combination. Occasionally an enzyme combination may produce repeat regions with the size selection window during subsequent batch processing.
- The quality of GBS data is influenced by your sample quality. The presence of degradation can cause false polymorphism detection in the analysis and create artificial bias in genotype calls. Please ensure adherence to our sample requirements (section 6).
- All samples will be quantitatively measured using QuantiFluor or the Qubit (high specificity dsDNA fluorometry). All samples will be visualised for quality on a 1% Agarose e-gel (Invitrogen), while the quality of a subset of samples will be further assessed on a GenomicTape (Agilent). AGRF assumes submitters have adhered to our quality guidelines and we do not bear responsibility for the performance of samples which were not qualitatively assessed by us.

3.0 Sequencing Chemistry

Batch processing libraries are sequenced on the NovaSeq 6000 platform using single read or paired-end 150 cycle sequencing.

4.0 Turn-around time

We provide a QC Report on all samples within a week of sample receipt for batches smaller than 288 samples. For the analysis, our 2 week TAT doesn't apply if the number of samples is more than 191.

Updates are provided at project milestones, but turn-around times may vary based on:
- Project sample set size
- Complexity of genome
- Bioinformatics requirements

Please discuss project design and turn-around time requirements with your Account Manager while quoting.

5.0 Sample Storage

Samples are stored with AGRF for 6 months after you receive your data. If you require your samples to be returned to you post-processing, please let your Account Manager know at the time of quoting. Please note that a fee will be charged for return of samples. If required, we can return samples on dry ice. Please note that dry ice shipments incur a $160 minimum (to metropolitan areas) shipping charge.

6.0 Sample Submission Requirements

6.1 Samples Requiring Extraction

Online Submission

- In the client portal, select 'Extraction and Genotyping by Sequencing' from the service dropdown menu
- Enter your species, tissue type and submission format (tube or plate)
- Complete and upload the template file
- Submit the form and print the submission receipt, and send with your samples to the addresses below:

Physical address (courier):
AGRF Adelaide
PLANT GENOMICS CENTRE
UNIVERSITY OF ADELAIDE
HARTLEY GROVE URRBRAE SA 5064

Postal address (mail):
AGRF Adelaide
PMB1 GLEN OSMOND
URRBRAE SA 5064

6.2 Extracted Samples (Purified DNA)

Online Submission

- In the client portal, select 'Genotyping by Sequencing' from the service dropdown menu
- Enter your species and submission format (tube or plate)
- Complete and upload the template file
- Submit the form and print the submission receipt to be included with your sample package
- NOTE: if your project includes an establishment phase, a separate online submission is required for these samples

DNA Requirements
Standard Establishment
Provide DNA from 3 individuals representative of your species:

- Buffer – HPLC Water
- Volume – 25µL
- Concentration – 20ng/µL (must be greater than 5ng/µL)

Batch Processing:

- Buffer – HPLC Water
- Volume – 20µL
- Concentration – 10ng/µL (must be greater than 5ng/µL)

Submission Format:

- Establishment samples – 1.5mL tube
- Batch processing samples – 96 well plate

Plate Format Submission Requirements:
- Leave a single well blank for AGRF negative controls
- Array samples down the plate
- Ensure each plate is heat-sealed or strip caps are used to prevent leakage in transit
- To prevent crushing during shipping, package plate/s in bubble wrap
- Label your plate/s with your Name, Contract ID and the plate number

Tube Format Submission Requirements:
- Please use 1.5mL snap cap tubes or screw cap tubes
- Please use parafilm to prevent tubes leakage in transit
- To prevent crushing during shipping, package tubes in bubble wrap

Sample Quality:
- DNA should be free from contaminants and extracted from pathogen- and symbiont-free tissues using the same method for all samples
- DNA should have an A260/280 ratio of 1.8-2 and an A260/230 ratio of >1.6
- We recommend that DNA concentration be assessed by fluorometry
- DNA should be of high molecular weight (>20 kb), free of RNA and not exhibit degradation when assessed by gel electrophoresis
- Samples with low DNA concentration should be concentrated by Zymo Research, gDNA Clean & Concentrator or AMPure XP beads by adjusting elution volume
- DNA should not be viscous during pipetting. Such samples may be less accessible to enzymes, reducing their contribution to final library preparation

6.3 Packing of Samples
- Samples can be shipped at room temperature via express post or courier
- To prevent leakage in transit please use parafilm to seal tubes leaking in transit, and ensure plates are heat-sealed or sealed with strip caps

Shipping Address:
GENOTYPING TEAM
AUSTRALIAN GENOME RESEARCH FACILITY
LEVEL 6, MRF BUILDING
ROYAL PERTH HOSPITAL
REAR 50 MURRAY STREET,
PERTH WA, 6000

If you have any questions regarding your sample submission please email CustomerCare@agrf.org.au.

7.0 Data Outputs
After sequencing is complete, your data will undergo analysis through our GBS analysis pipeline. Our pipeline uses Stacks software version 2.60 (http://catchenlab.life.illinois.edu/stacks/). The software demultiplexes the sequencing reads and creates a separate FASTQ file for each sample. Our standard analysis is based on a *de novo* alignment. The software creates stacks of similar reads for each sample individually. These read stacks are also known as tags. The tags which appear across all samples are collated (catalogue tags), and genotypes are calculated for the common polymorphic sites. If a reference is available and a reference-based analysis is preferred, please speak with your account manager.

Analysed data is provided in .vcf, .fastq and .txt formats, and raw data is provided in .fastq.gz format. A project report is also provided which summarises key statistics for each sample such as the number of catalogue tags, the average tag depth and read number.

Custom analyses can be added to your project upon request. If you have specific analysis requirements please discuss these during the quotation process. Addition of custom analysis components will incur additional expense and cannot be offered within our standard turnaround time. An expected turnaround time will be communicated during the quotation process.

If you do not wish to include AGRF bioinformatics analysis in your project, raw data will be provided in .fastq.gz format as well as a text file to assist in demultiplexing.