**BASE Project Workflow**

**Thorough documentation of the collection procedures will ensure maintenance of sample value.**

1. Following acceptance of your samples by the selection group: Request unique identifiers for sample naming/labelling from Bioplatforms (Anna Fitzgerald email: [afitzgerald@bioplatforms.com](mailto:afitzgerald@bioplatforms.com))
2. Select a 25 x25m plot at the sample site in a reasonably homogenous environment that reflects the characteristics of the site (based on soil, vegetation and land use).
3. Collect soil (comprising between 9-30 samples) from the plot in a manner that adequately samples the whole plot and ensures biological integrity of the sample.

These can be sampled regularly on a grid if you think there isn’t much microgeographic variation, or the sample points could be stratified to take account of anything you think might be important.

1. Collect the samples as two depths (1) top 10cm; (2) 20cm and below (define).

Homogenise all within plot sub-samples to make a pooled sample for each of the two depths. Samples are best mixed in the field in a large ziplock plastic bags from which aliquots can easily be drawn. Sieve if needed.

1. Ensure collection of adequate soil for nucleic acid extraction and contextual data generation from each sampling unit (e.g., 1 kg from each depth).
2. Assign unique sample identifiers to each sample using the last digits after the “/” i.e. 102.100.100/8202 would be called **8202**. Each depth is to have a separate identifier.
3. Collect and record all other local contextual data listed in the ‘BASE contextual data template spreadsheet’.
4. Take photos of plot – soil and surrounding environment (example below).



1. Send each sample for DNA analysis, chemical analysis and archiving as below:
2. **DNA ANALYSIS**:

* Fill 50mL Falcon tube with soil for DNA analysis (leaving 1-2 cm of space at top of tube) and freeze as soon as possible.
* Contact John Stephen from AGRF at the address below to organise

(1) soil transfer permit paperwork (Biosecurity SA)

(2) ensure you have a current login for the AGRF sample submission site

(3) log samples online via the AGRF website (follow instructions below).

* Send frozen samples to AGRF Adelaide for DNA extraction and sequencing analysis.

|  |  |  |
| --- | --- | --- |
| **Facility** | **Address** | **Contact person** |
| **AGRF** | Australian Genome Research Facility Ltd  Plant Genomics Centre  University of Adelaide  PMB1 Glen Osmond  Urrbrae SA 5064 | John Stephen  Phone: (08) 8303 7202  Email: [John.Stephen@agrf.org.au](mailto:John.Stephen@agrf.org.au) |

Sample submission instructions:

1. **Submitter Registration**

* All work is being performed under a single agreement in the AGRF system:

Agreement ID: AGRF12-00265

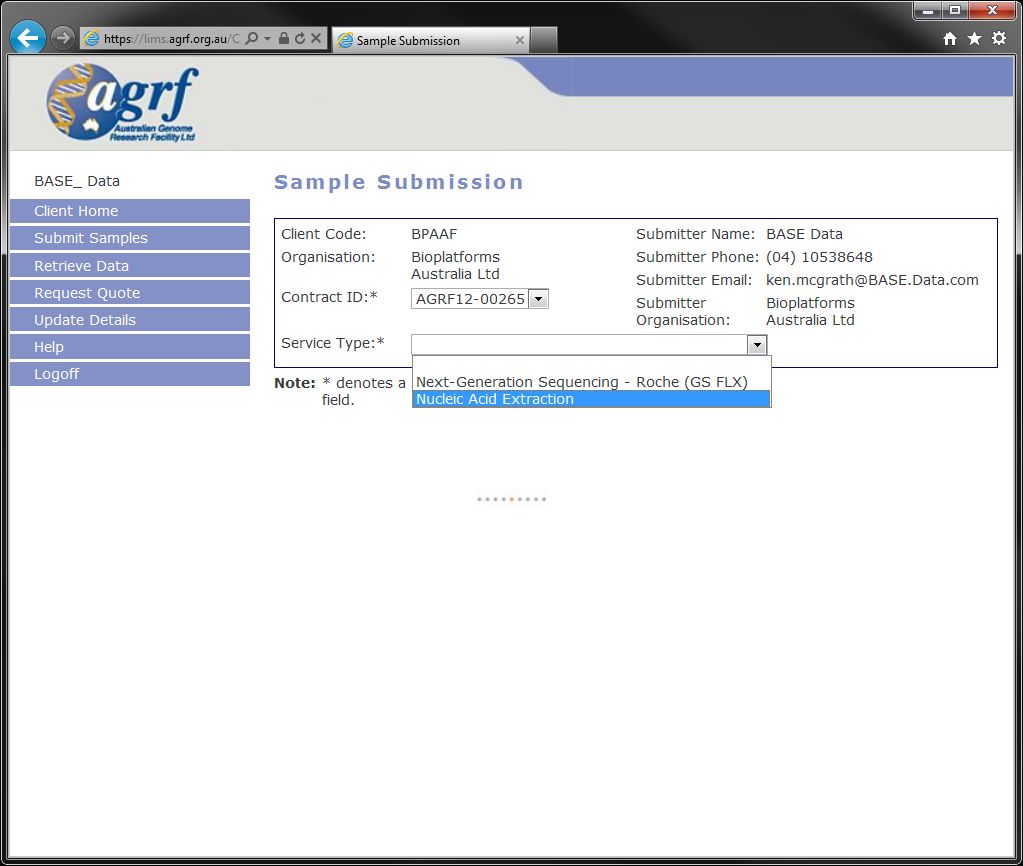
Primary contact: Anna Fitzgerald

* Register to this account on the AGRF website <https://lims.agrf.org.au/ClientSubmission/NewUser.aspx>
* Tick the option that says “I want to register as a sample submitter on an existing contract/account”, and use contract ID number “AGRF12-00265”
* An email will be sent to Anna Fitzgerald, requesting authorisation. Submissions can be made online immediately, but sample processing will not occur until authorisation has been received.

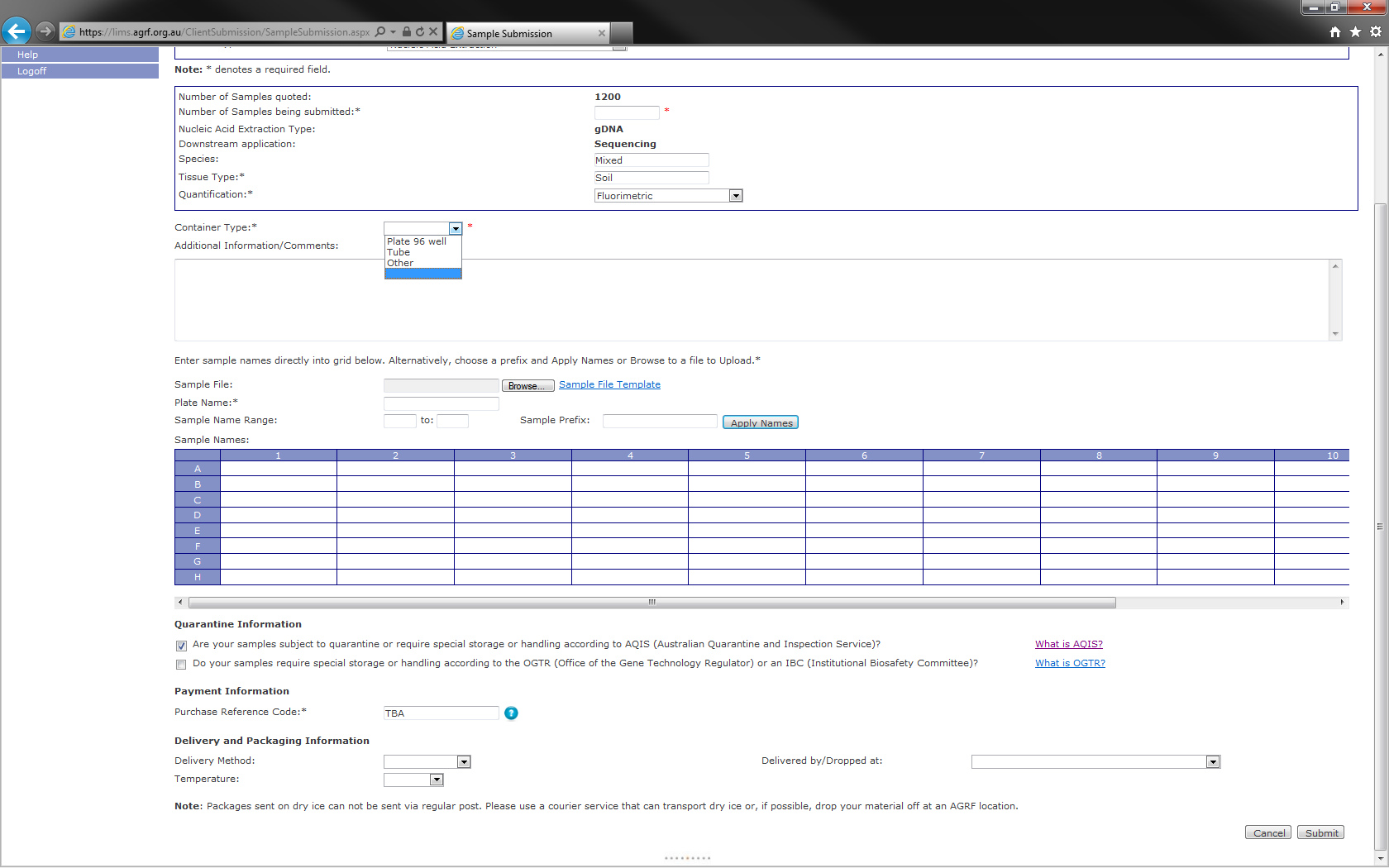
1. **Submission process**

Once registered:

* log into the AGRF website by clicking the “update details” link on the left hand side. Enter the contract ID AGRF12-00265 and click submit.
* Then click on the “submit samples” link on the left side menu.
* Choose the contract ID “AGRF12-00265”, and then select the “Nucleic Acid Extraction” option (see image below).



* Once selected, a page asking for submission details will appear:



* Complete the relevant fields, and click submit.
* A link will appear at the top of the page to a PDF document; this your submission receipt.
* Print off the sample submission receipt, and include this with your samples that you send to Adelaide.
* Samples are to be shipped frozen (dry ice). Any samples received that are thawed/room temperature when inspected will be deemed inappropriate for extraction.

Gently air dry and bag remaining soil for:

1. **CHEMICAL ANALYSIS (as below or equivalent)**:

* Send sample 250g dry weight preferred [200g minimum required] to CSBP, Perth WA for the ‘Comprehensive Test’ .

<http://www.csbp-fertilisers.com.au/nutrition-services/soil-and-plant-analysis-laboratory>

* Cost per sample will be $52.50+GST when citing quote number: CSBP13/0457 and account number 55877. Please include your “bill to” contact and address.
* Include completed ‘Soil Analysis Request Form’ and ‘Payment Information’ form available as pdfs from the CSBP website.

|  |  |  |
| --- | --- | --- |
| **Facility** | **Address** | **Contact person** |
| **CSBP** | CSBP Soil and Plant Laboratory  2 Altona Street  BIBRA LAKE WA 6163 | Phone: (08) 9434 4600  Email: analysis@csbp.com.au |

1. **ARCHIVING**:

* Remaining soil to be sent to CSIRO National Soils Archive, Canberra.
* Samples must be air dried (or oven dried at 40 degrees Celsius)
* Minimum 200g dry weight to be sent in approved containers (Cospak 500mL or 1L polyethylene containers cat # A500 or A572) labelled with BPA provided unique identifier.
* Email Linda Karssies at the address below when you are sending samples for archive so they are aware of their impending arrival.

|  |  |  |
| --- | --- | --- |
| **Facility** | **Address** | **Contact person** |
| **CSIRO National Soils Archive** | National Soils Archive  CSIRO Land and Water,  Black Mountain,  Canberra ACT 2601 | Linda Karssies  Phone: (02) 6246 5824  Email: [linda.karssies@csiro.au](mailto:linda.karssies@csiro.au) |

1. **SITE CONTEXTUAL DATA:**
2. Submit all contextual data into the excel flat file sheet supplied by Anna Fitzgerald.

**DNA EXTRACTION WORKFLOW (AGRF Adelaide):**

1. DNA extracted from each sample in triplicate using the Mo Bio PowerLyser Power Soil DNA extraction kit (250mg input into each extraction).

* For consistency if added DNA output is required an increased number of extractions will be performed and pooled and noted in the metadata (rather than increasing the soil input into an extraction)

1. DNA replicate extractions pooled.
2. DNA extractions will be named using the last 4-5 digits of the soil sample unique identifier followed by an underscore and the extraction number e.g. sample 102.100.100/8202 would be called **8202\_1** and if a repeat extraction is required at a later date the DNA sample would be labelled **8202\_2** etc
3. DNA sample QC:

* Fluorimetric quantification of DNA measured by picogreen fluorometry (samples must have greater than 1ng/ul to continue in workflow)
* Qualitative PCR assay for polymerase inhibition performed on the extracted samples (16s – V3-V8 region):
* Samples that do not inhibit amplification will be deemed appropriate for further processing.
* Samples with PCR inhibiting activity will be diluted 1:100 and retested. If that dilution no longer inhibits PCR, it will be forwarded for further processing.
* If still inhibitory after 1:100 sample will be put aside for further troubleshooting (e.g. purification on the Aurora platform).
* Samples that don’t pass QC after Aurora purification will be re-extracted once more with option to increase number of reactions if from a low biomass area – if fail a second time samples will be put aside for trouble shooting at CSIRO
* Samples undergoing metagenomic analysis to be cleaned using zymo concentrate columns
* QC results to be entered into Google doc

<https://docs.google.com/spreadsheet/ccc?key=0AvChXk8QAupBdGVxYmNXZjM4TG9yS2FmRXhQWkUtNFE&usp=drive_web#gid=0>

1. Samples that pass QC to be sent to sequencing facilities and EMP (only samples from top 0-10cm to be sent to EMP).
2. Aliquot (50 µL) taken for Earth Microbiome Project (EMP) Illumina analysis (sent when DNA available from the total pool of samples requested by the EMP).
3. Samples (~20ul) to be shipped in batches of 93 to AGRF Melbourne or Ramaciotti Centre, UNSW

(leaving room for negative control, mock community positive control, BASE soil sample community positive control)

**AMPLICON SEQUENCING WORKFLOW:**

To be undertaken at AGRF Melbourne and Ramaciotti Centre, UNSW

1. **Amplicon generation**

Two positive controls to be included with each run:

1. mock community positive control (Bacterial for 16s and fungal for ITS and 18s)

ii) BASE soil sample community positive control (pooled DNA extraction from samples as below – supplied at 25ng/ul)

|  |
| --- |
| Samples included in BASE soil sample community positive control |
| 7061  9577  8511  8487  8281  8196  7859  9507  7896  2409 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Amplicon gene target** | **Target organism** | **Target region** | **Primers** | **Technology Platform** | **Current estimated # of reads** |
| 16S | Bacteria | V1-3 | 27F/519R | Illumina Miseq  2x300 bp paired end reads | ~200,000 per sample |
| Bacteria and archea | V4 | 515F/806R | **EMP collaboration**  Illuminia HiSeq2500 150bp paired end reads | - |
| Archaea | V3-V5 | Arch-340F/Arch-915R1 | Illumina Miseq  2x300 bp paired end reads | ~200,000 per sample |
| ITS | Fungi | ITS1-4 | ITS1F/ITS4 | Illumina Miseq  2x300 bp paired end reads | ~200,000 per sample |
| 18S | Euk/algal | V9 | Euk1391f/ EukB | Illumina Miseq  2x150 bp paired end reads | ~100,000-140,000 per sample |

1Pinto AJ, Raskin L (2012) PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. PLoS ONE 7(8): e43093.

1. **Amplicon library preparation and QC:**

* Sample input 1:10 dilution. If fails reflex input to neat or 1:100 depending on DNA concentration.
* Bead clean up preformed on PCR product.
* Quantify PCR product by Quant-iT DNA or PicoGreen Fluorimetry.
* Verify PCR product size by Bioanalyser or Labchip.
* If can’t determine size or concentration too low to detect on labchip or Bioanalyser sample will be disregarded.

1. **Amplicon sequencing:**

* SOP to be included as an attachment for each target
* Uniform MiSeq file naming format:

Use excel template to concatenate the BPA ID, target, facility, flow cell ID and index sequence to produce a file name consistent with the HiSeq (as in table below).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **BPA ID** | **Target** | **Facility** | **Index** | **Flow Cell** | **Sample Name** |
| XXXX\_1 | 16S | AGRF | GGAGACAAGGGA | A5BHR | XXXX\_1\_16S\_AGRF\_GGAGACAAGGGA\_A5BHR |
| XXXY\_1 | 16S | AGRF | AATCAGTCTCGT | A5BHR | XXXY\_1\_16S\_AGRF\_AATCAGTCTCGT\_A5BHR |
| XXXZ\_1 | 16S | AGRF | AATCCGTACAGC | A5BHR | XXXZ\_1\_16S\_AGRF\_AATCCGTACAGC\_A5BHR |

**Target code**

* 16S
* 18S
* ITS
* A16S

**Facility code**

* UNSW
* AGRF

Final file name: **SampleName \_S#\_L001\_R1\_001.fastq.gz**

SampleName = in Sample\_Name column (as in table above)

S# = sample number in the run

L001 = lane 1

R1 = Read 1 or R2 = Read 2

e.g.

7709\_1\_16s\_AGRF\_GGAGACAAGGGA\_A5BHR \_ S#\_L001\_R1\_001  
8201\_1\_18S\_UNSW\_AATCAGTCTCGT\_A5BHR S#\_L001\_R1\_001

Consistency in naming – UNSW, AGRF16S, 18S, ITS and index sequence should always be upper case.

1. **Amplicon post run processing:**

* Use default machine settings of Miseq software version information to be included with sequencing metadata
* Machine set up for fastq index file output in addition to read 1 and read 2 compressed fastq file outputs
* Forward, reverse and index read files with checksums to be sent for data storage

**METAGENOMICS SEQUENCING WORKFLOW:**

To be undertaken at AGRF Melbourne and Ramaciotti Centre, UNSW

1. **Metagenomics library preparation and sequencing:**

* Samples prepared using Illumina TruSeq PCR Free kit (350 or 550bp insert size) as per Illumina instructions
* Libraries to be sequenced on Illumina HiSeq 2500 in rapid mode generating 0.5 lanes of 150bp PE data per sample (final data volumes to be confirmed)
* Uniform HiSeq file naming format:

**UniqueID\_extraction\_BASE\_sequencing centre code\_FlowID\_Index\_Lane\_F1/R1Read**

e.g. Correct file naming

7709\_1\_BASE\_UNSW\_DOVOEACXX\_NoIndex\_L001\_R1

**Facility code**

* UNSW
* AGRF

Consistency in naming – UNSW, AGRF, flow cell ID and index sequence should always be upper case; and if no index used, please label NoIndex, not NOINDEX.

1. **Metagenomics post-run processing:**

* Demulitplexing with CASAVA software
* Compressed fastq files with checksums to be sent for data storage.

**DATA MANAGEMENT AND STORAGE WORKFLOW:**

1. **Sequencing metadata:**

Sequencing facilities to complete spreadsheet with all requested metadata

1. Amplicon: “BASE\_Amplicon\_Sequencing\_Metadata\_Sheet” excel spreadsheet to be completed for each set of data and emailed to CCG at time of file transfer
2. Metagenomics: <https://docs.google.com/spreadsheet/ccc?key=0AvChXk8QAupBdENpekdlWGQxcnM1dHhMOGVRVVBGdXc&usp=sharing#gid=0>
3. **Data transfer:**

* Data and QC information for successful sample runs (reaching the required number of reads only) to uploaded by sequencing facilities to CCG for storage and controlled access (through iVEC).
* A new directory is to be made for each new data set uploaded. New directories to be labelled with a date stamp e.g. 20131216\_AGRF\_BPA\_BASE\_D251JACXX.
* Directory to include:

1. All sequencing data files
2. Sequencing metadata for all files being transferred in Excel spreadsheets – for amplicon data “BASE\_amplicon\_sequencing\_metadata\_template\_V1” or metagenomics data “BASE\_metagenomics\_sequencing\_metadata\_template\_V1”
3. File names and and corresponding MD5 checksums information in a separate text file for all files being transferred.

* Email to be sent to CCG (both Thys and Matt) when data is made available with the following information:
* Project the data relates to (i.e. BASE project)
* Name of relevant date stamped file where data is residing

|  |  |  |
| --- | --- | --- |
| **Facility** | **Address** | **Contact person** |
| **Centre for Comparative Genomics** | Thys Meintjes  Centre for Comparative Genomics  Room 3.002, BITL Building  Murdoch University  Perth, WA 6150 | Thys Meintjes  Phone: (08) 9360 2422  Email: [tmeintjes@ccg.murdoch.edu.au](mailto:tmeintjes@ccg.murdoch.edu.au)  Matt Bellgard  Email: [mbellgard@ccg.murdoch.edu.au](mailto:mbellgard@ccg.murdoch.edu.au) |

1. CCG to notify appropriate data user of data availability and details for FTP access .
2. Amplicon data will be processed through the QIIME software pipeline.
3. Raw and processed data will be made available through the Bioplatforms Australia Downloads page: <https://downloads.bioplatforms.com/>



