

## 775 Import Format Layout

### **SNP Data**

Result must contain calls and no-calls.

There are three sections included in the 775 format:

- File Header – occurs once at the start of the file
- Test Header – occurs once for each animal in the file
- Test Data – contains the SNP data results for each animal, one record for each SNP

#### **[FILE-HEADER] Section:**

- There is one FILE-HEADER section per file
- Field names are uppercase and include trailing colon
- Field name and field value are separated by a tab
- We are on Version 3 of this format.

<b>Field Name</b>	<b>Field Value</b>	<b>Optional or Required</b>
FORMAT-VERSION:	3	Required
LAB-ID:	DNA Lab Identifier	Required

*A complete list of DNA Lab Identifiers is available on the BREEDPLAN Web Site:  
([http://breedplan.une.edu.au/index.php?option=com\\_content&task=view&id=71](http://breedplan.une.edu.au/index.php?option=com_content&task=view&id=71))*

[TEST-HEADER] Section:

- There is a TEST-HEADER section usually with a corresponding TEST-DATA section for each animal's test results.
- Field names are uppercase and include trailing colon
- Field name and field value are separated by a tab

Field Name	Field Value	Optional or Required
GSGT-VERSION:	Version number	Optional
PROCESSING-DATE:	YYYYMMDD HHMMSS (if supplied time in 24H format)	Date is required Time is optional
CONTENT:	SNP chip description (Note that this will be used to link to a SNP Map with this Identifier)	Required
CALL-RATE:	Proportion of NUM-SNPS to TOTAL-SNPS (eg 0.9601)	Optional
TOTAL-SNPS:	Total number of SNPs on the SNP chip	Required
NUM-SNPS:	Number of SNPS in the [TEST-DATA] section for this animal	Required
REQUEST-TYPE:	Request type code sent with request eg. GGPHD	Required
BATCH-NO:	Batch number as provided in request.	Required if provided in request
SAMPLE-NO:	Sample number within the batch as provided in request.	Required if provided in request
SENTRIX-POSITION:	Sentrix position (sample, location, plate).	Optional
DNA-CASE-ID:	DNA-case-id (unique identifier allocated to the animal and its sample material by DNA Lab)	Required
ANIMAL-ID:	The animal ident exactly as recorded on the society database. This is supplied in the request if you have been issues an automated ILR2 request.	Required

**[TEST-DATA] Section:**

This Data section should include the following field values:

<b>Order</b>	<b>Field Description</b>	<b>Optional or Required</b>
1	SNP Name	Required – uppercase
2	Allele1 - Forward	Optional
3	Allele2 - Forward	Optional
4	Allele1 - Top	Optional
5	Allele2 - Top	Optional
6	Allele1 - AB	Required
7	Allele2 - AB	Required
8	GC Score	Required
9	X	Optional
10	Y	Optional

- If no SNP results were obtained for the animal, please only include the TEST-HEADER section and ensure NUM-SNPS = 0. In this case, there will be no TEST-DATA for the animal.
- If no field value is available then leave field blank, but include all field separators (tabs)
- If a particular SNP test did not give a result, please include the SNP Name and leave other fields blank. (No-Calls)



## SNP Data File Format – V3 For loading via ILR2 775 Format

Last Updated March 2019

### Example File

```
[FILE-HEADER]
FORMAT-VERSION: 1
LAB-ID: AUUQLD
[TEST-HEADER]
GSGT-VERSION: 1.9.4
PROCESSING-DATE: 20120305 1354
CONTENT: BOVINESNP50_V2_C.BPM
TOTAL-SNPS: 57501
NUM-SNPS: 54609
REQUEST-TYPE: 50K
BATCH-NO: 1210095
SAMPLE-NO: 1
SENTRIX-POSITION: 200693460002-R08C01
DNA-CASE-ID: 1299991
ANIMAL-ID: ABCE10
[TEST-DATA]
ARS-BFGL-BAC-10245 T C A G A B 0.7579 1.317 0.939
ARS-BFGL-BAC-10375 G G G G B B 0.9567 0.006 0.716
ARS-BFGL-BAC-10591 G G G G B B 0.9003 0.017 0.672
ARS-BFGL-BAC-10867 G G C C A A 0.9434 0.513 0
ARS-BFGL-BAC-10919 A G A G A B 0.8526 0.698 0.783
[TEST-HEADER]
GSGT-VERSION: 1.9.4
PROCESSING-DATE: 20120305 1400
CONTENT: BOVINESNP50_V2_C.BPM
TOTAL-SNPS: 57501
NUM-SNPS: 54609
REQUEST-TYPE: 50K
BATCH-NO: 1210095
SAMPLE-NO: 2
SENTRIX-POSITION: 200693460002-R06C01
ANIMAL-ID: ABCE11
DNA-CASE-ID: 1299992
[TEST-DATA]
ARS-BFGL-BAC-10245 T C A G A B 0.7579 1.317 0.939
ARS-BFGL-BAC-10375 G G G G B B 0.9567 0.006 0.716
ARS-BFGL-BAC-10591 G G G G B B 0.9003 0.017 0.672
ARS-BFGL-BAC-10632
ARS-BFGL-BAC-10867 G G C C A A 0.9434 0.513 0
ARS-BFGL-BAC-10919 A G A G A B 0.8526 0.698 0.783
```

The SNP data file name should be

**<SOC\_CODE>\_<DNA\_LAB\_ID>\_775\_<BATCH\_NO>\_yyyymmdd\_hhmm\_SNP\_DATA.TXT**

**SNP\_Map.txt** (tab delimited text file). This file is Compulsory for all genotypes except parentage.

The SNP Map is important for our quality control processes. If it is not supplied, an error will occur, and the result data will not be loaded onto ILR2. The SNP Names listed in the SNP Map should correspond exactly to the SNP Names supplied within the 775 file, including the 'No Calls'.

Field Name	Field Description	Optional or Required	Example
Index	Row sequence no	Required	26
Name	SNP Name	Required	ARS-BFGL-BAC-10919
Chromosome	Chromosome Number	Required	14
Position		Required	31267746
GenTrain Score		Optional	0.7455
SNP		Required	[A/G]
ILMN Strand	Top or Bottom	Optional	TOP or BOT
Customer Stand	Top or Bottom	Optional	TOP or BOT
NormID		Optional	100

### Example File

```

1 |Index |Name |Chromosome |Position |GenTrain Score |SNP |ILMN Strand |Customer Strand |NormID|CR103
2 |1 |ARS-BFGL-BAC-10919 |14 |31267746 |0.7455 |[A/G] |TOP |TOP |0|CR103
3 |2 |ARS-BFGL-BAC-10975 |10 |21225382 |0.7042 |[A/G] |TOP |TOP |0|CR103
4 |3 |ARS-BFGL-BAC-11000 |10 |79252023 |0.8459 |[T/G] |BOT |BOT |0|CR103
5 |4 |ARS-BFGL-BAC-11003 |10 |80410977 |0.8801 |[T/C] |BOT |BOT |0|CR103
6 |5 |ARS-BFGL-BAC-11025 |10 |84516867 |0.8560 |[T/G] |BOT |BOT |0|CR103
7 |6 |ARS-BFGL-BAC-11044 |1 |12805406 |0.8861 |[T/C] |BOT |BOT |0|CR103
8 |7 |ARS-BFGL-BAC-11193 |1 |29303546 |0.8123 |[T/C] |BOT |TOP |0|CR103
9 |8 |ARS-BFGL-BAC-11215 |12 |90704572 |0.7441 |[A/G] |TOP |TOP |0|CR103
10 |9 |ARS-BFGL-BAC-11218 |1 |24549757 |0.8716 |[A/G] |TOP |TOP |0|CR103
11 |10 |ARS-BFGL-BAC-11276 |13 |2153905 |0.8155 |[T/C] |BOT |TOP |0|CR103
12 |11 |ARS-BFGL-BAC-11283 |13 |23196342 |0.8106 |[A/C] |TOP |TOP |0|CR103
13 |12 |ARS-BFGL-BAC-11513 |14 |7428315 |0.7885 |[A/G] |TOP |TOP |0|CR103
14 |13 |ARS-BFGL-BAC-11612 |10 |27228918 |0.8355 |[T/C] |BOT |BOT |0|CR103
15 |14 |ARS-BFGL-BAC-11657 |10 |48926752 |0.8042 |[T/C] |BOT |TOP |0|CR103
16 |15 |ARS-BFGL-BAC-11666 |10 |49656226 |0.8151 |[A/G] |TOP |TOP |0|CR103
17 |16 |ARS-BFGL-BAC-11718 |0 |0 |0.6882 |[T/C] |BOT |BOT |0|CR103
18 |17 |ARS-BFGL-BAC-11724 |11 |60498991 |0.6985 |[A/G] |TOP |BOT |0|CR103
19 |18 |ARS-BFGL-BAC-11748 |11 |58442169 |0.9014 |[G/C] |BOT |TOP |100|CR103
20 |19 |ARS-BFGL-BAC-11750 |1 |19778380 |0.8753 |[T/C] |BOT |TOP |0|CR103
21 |20 |ARS-BFGL-BAC-11761 |1 |19709807 |0.7892 |[A/G] |TOP |TOP |0|CR103

```

SNP Map **must** accompany the SNP DATA. The file naming convention is:  
**<SOC\_CODE>\_<DNA\_LAB\_ID>\_788\_<BATCH\_NO>\_yyyymmdd\_hhmm\_SNP\_MAP.txt**

This SNP MAP file will be used to validate against the result in the 775 file.

**AnimalDetails.csv** (comma separated csv file). This file is optional (but can be beneficial).

This file allows us to determine which result belongs to which animal. It can be used as a cross reference from the data in the 775 result to an actual animal registered on the society database via the 'Barcode Id' in the AnimalDetails matching the DNA-CASE -ID in the SNP DATA file. However, its only needed if the Ident in the SNP DATA file is NOT the real registered animal identifier.

Field Name	Field Value	Optional or Required	Example
Order	Order number allocated by the society or the laboratory. Often Order & Batch ID are the same value.	Optional	123456
Batch ID	A batch number allocated by the society or the laboratory. Often Order & Batch ID are the same value.	Optional. Beneficial If the society has allocated via ILR2	1901001
Barcode	Either the DNA Case Id or the TSU #	Required	654321
Call Rate	Proportion of the NUM-SNPS to TOTAL-SNPS (eg.09982). The quality of the result.	Optional	0.9982
File ID	Name of the results file.	Optional	20180308_CHAR_multibatch_BOVG50V01
Animal ID	The animal ident exactly as recorded on the society database, or the Animal Identifier as provided in the request.	Required	XXXY9999
Registration Number	The animal ident exactly as recorded on the society database ie the society allocated Identifier for a recorded animal.	Optional. Required if known	XXXY9999
Ear Notch	Notch details as provided in the request	Optional	
Barcode 2	The DNA Case ID if the TSU# is in Barcode otherwise DNA Case ID.	Required if Barcode set to TSU#	654321
Sample Type	Described the sample type	Optional	Hair, Semen, TSU
Date of Birth	dd/mm/yyyy	Optional	15/04/2019
Breed	Breed Description	Optional	CHAROLIAS
Sex		Optional	M or F
Organization	Full name of the society the animal or request belongs to	Optional	CHAROLAIS SOCIETY OF AUSTRALIA LTD
Member Code	The Membership ident allocated to the member by the society.	Optional	XXX

## Data Example

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Order	Batch ID	Barcode	Call Rate	File ID	Animal ID	Registration Number	Ear Notch	Barcode 2	Sample Type	Date of Birth	Breed	Sex	Organization	Member Code	
2	354248	354248	672326	0.9982	20180308_CHAR_multibatch_BOVG50V01	DAS N13E	DAS N13E		672326	Hair		CHAROLAIS	M	CHAROLAIS SOCIETY OF AUSTRALIA LTD	DAS	
3	354248	354248	672327	0.9982	20180308_CHAR_multibatch_BOVG50V01	DAS N15E	DAS N15E		672327	Hair		CHAROLAIS	M	CHAROLAIS SOCIETY OF AUSTRALIA LTD	DAS	
4	354248	354248	672328	0.9982	20180308_CHAR_multibatch_BOVG50V01	DAS N2E	DAS N2E		672328	Hair		CHAROLAIS	M	CHAROLAIS SOCIETY OF AUSTRALIA LTD	DAS	
5	354248	354248	672329	0.9984	20180308_CHAR_multibatch_BOVG50V01	DAS N12E	DAS N12E		672329	Hair		CHAROLAIS	M	CHAROLAIS SOCIETY OF AUSTRALIA LTD	DAS	
6	354249	1802007	672330	0.9983	20180308_CHAR_multibatch_BOVG50V01	DOULTON LARRIKIN (DSM L4E) (P)	DSM L4E		672330	Hair	02/10/2015	CHAROLAIS	M	CHAROLAIS SOCIETY OF AUSTRALIA LTD	DSM	

The Animal details file is an extra file of animal details may be delivered with the SNP data file. This file contains extra details of the animals included in the SNP\_DATA file.

The file naming convention will be:

**<SOC\_CODE>\_<DNA\_LAB\_ID>\_787\_<BATCH\_NO>\_yyyymmdd\_hhmm\_AnimalDetails.CSV**

## File Naming Convention

**<SOC\_CODE>\_<DNA\_LAB\_ID>\_775\_<BATCH\_NO>\_yyyymmdd\_hhmm\_SNP\_DATA.zip**

where <SOC\_CODE> is the ABRI assigned Society Code \*

<DNA\_LAB\_ID> is the DNA Lab Identifier #

<BATCH\_NO> is the lab requests batch number. If no request has come from the society then enter 7 zeros.

yyyymmdd\_hhmm is a date and time stamp (24 hr format).

\* ( [http://breedplan.une.edu.au/booklets/Soc\\_codes\\_BP.pdf](http://breedplan.une.edu.au/booklets/Soc_codes_BP.pdf))

# ( <http://breedplan.une.edu.au/booklets/DNA%20Genetic%20Labs%20Codes.pdf>)

Example: filename for batch 1074 from The University of Queensland created at 13:12:01 on the 9<sup>th</sup> Nov 2016 for animals in the Australian Wagyu Association:

AUWY\_AUUQLD\_775\_1074\_20161109\_1312\_SNP\_DATA.TXT

## ZIP File

The .....SNP\_DATA , the .....SNP\_MAP.txt and the .....AnimalDetails.csv (if available) must be compressed into a single ZIP file. The file naming convention for the ZIP file will be:

**<SOC\_CODE>\_<DNA\_LAB\_ID>\_775\_<BATCH\_NO>\_yyyymmdd\_hhmm.ZIP**

Example filename for batch 1074 from The University of Queensland created at 13:12:01 on the 9<sup>th</sup> Nov 2016 for animals in the Australian Wagyu Association:

AUWY\_AUUQLD\_775\_1074\_20161109\_1312.ZIP

## ILR2 (International Livestock Register v2)

ILR2 has an automated process where the zip file is loaded into the application and the system will unzip the file and process the files contained within. Therefore the naming convention is important for this automation to be successful.

## File Download Area

A single DNA data download area has been set up as a central repository and distribution point for all ABRI clients (except Aust Angus).

ABRI will:

- Check the files and advise the DNA Lab if there are any problems with file names.
- Automatically transfer the files to the relevant Breed Society ILR2 database.
- Notify the Breed Society that the file has been sent from the DNA Lab and that the Import process has been initiated.

Each DNA Lab will have its own username and password. Contact ABRI to get these details.

\* end \*