The Angus Society of Australia Ordering DNA Services

Export File Format

Table of Contents

Introduction

One part of the Angus Society of Australia's online system, https://angus.tech/, is the DNA module. The DNA module allows members to request DNA testing online. This is an alternative to filling in and mailing the paper form[1].

The user interface provides an easy way to provide animal, sample and test details. However, an appropriately-formatted CSV file can also be used. This document describes the format of this CSV file.

The purpose for accepting CSV files is twofold.

Firstly, *members who are more comfortable with Microsoft Excel can continue submitting details without change*. That is, those who previously filled in the template[2] from the society's website can continue to do so. They can then either email the spreadsheet in (as per the current processes) or they can convert the file to CSV (via the *Save As* menu item) and upload it to angus.tech themselves[3].

Secondly, *software vendors can easily produce a CSV file*. For example, a herd management program could produce a CSV requesting tests for *Developmental Duplications* for all calves born in the current year.

---

1   The PDF generated on submission, along with any physical samples, must still be mailed to The Angus Society of Australia.

2   The Microsoft Excel template is downloadable from the society's website at https://www.angusaustralia.com.au/registrations/dna/dna-forms/.

3   Older versions of the Microsoft Excel template are incompatible. Only versions of the template produced after the public launch of the DNA module are compatible.

A General Overview of CSV Files

This section describe CSV formats in general. For details specific to the CSV files acceptable in the DNA module, see the section *DNA Module CSV Files.*

A CSV file is a human-readable, two-dimensional grid of data. Each line of text contains a *row* separated into *fields* using commas.

The example CSV file below contains four rows, each with two fields.

```
ANIMAL_ID,SAMPLE_BARCODE
ABCX1,98200000001
ABCX2,98200000002
ABCX3,"982,00000003"
```

The first row contains two fields: "*ANIMAL_ID*" in field 1 and "*SAMPLE_BARCODE*" in field 2. Often, the first row contains headers, indicating how each field of subsequent rows should be interpreted. This allows you to, for example, change the order of fields.

```
SAMPLE_BARCODE,ANIMAL_ID
98200000001,ABCX1
98200000002,ABCX2
"982,00000003",ABCX3
```

If a field value contains a comma itself, it should be surrounded by double quotes. For example, the sample barcode for animal ABCX3 is "*982,00000003*" - it contains a comma and is 12 characters long.

The are many variations of the CSV file format. For example, some use the tab character (hexadecimal `0x09`) or the pipe character (`|`) to separate fields. There are also many scenarios which are not specified. For example, how the software should interpret fields that contains double quotes (`"`), carriage returns and new lines (hexadecimal `0x10` and `0x13`, respectively), or unterminated quoted fields.

In regards to the DNA module in angus.tech, spreadsheet software like *Microsoft Excel* generally produce acceptable CSV output. Also, given the types of values being produced, erroneous output should rarely occur and when it does, it can be easily corrected.

DNA Module CSV Files

Types of CSV Files

Two types of CSV files can be uploaded:

- The *Animals, Samples and Tests* CSV file

- The *Parent Verifications* CSV file

Both types of CSV files follow the "row *type*" formatting pattern.

The Row Type Formatting Pattern

The *"row type"* formatting pattern uses the first field of each row of a CSV file to indicate what should be done with the remaining fields of that row. In both types of CSV files in the DNA module, there are three row types.

The first row type is the *ignored* row type. If a value of "IGNORE" appears in the first field, that row is discarded.

The second row type is the *header* row type. If a value of "HEADER" appears in the first field of a row, the remaining fields of that row are treated as headers. As in the above section ("*A General Overview of CSV Files*"), the header indicates how values should be interpreted and allows you to change the order of fields in a CSV file.

The third row type is the *data* row type. If an empty string appears in the first field of a row, the remaining fields of that row are treated as values. The empty string was chosen rather than a more-descriptive string (e.g., "DATA") so that users can insert additional rows using spreadsheet software like Microsoft Excel without being aware of the "row type" concept.

Note that there are additional rules when processing CSV files. For example:

- Data rows where every field is empty are ignored.

- Unknown row type are treated as errors.

- There can only be one *header* row type.

- The *header* row type must appear before all *data* row types.

The Animals, Samples and Tests CSV File

Overview

As the name suggests, this file contains details on the animals, samples and tests that are part of the order. That is, it describes the samples that should be tested, the animals to which those samples belong, and the tests to be performed on each sample.

Fixed Fields

This CSV requires five fields which must always be present.

The first field is the "SAMPLE_TYPE" field, which contains the type of sample being supplied. Valid values are "H" for hair, "T" for tissue, "U" for tissue contained in TSU collectors, "S" for semen and "E" for existing. Note that if the sample type is not "E" for existing, the sample is expected to be a newly supplied sample. Any value not specified above is an error.

The second field is the "SAMPLE_BARCODE" field, which contains the barcode that appears on the sample. Note that for newly supplied samples (where "SAMPLE_TYPE" is "H", "T", etc) this field is required. For existing samples (where "SAMPLE_TYPE" is "E") this field is optional. If specified, it must contain the barcode of the sample to be used. If not specified, the system will choose an appropriate existing sample (e.g., a recent sample that is suitable for testing).

The third field is the "COLLECTOR_TYPE" field, which identifies the type of collector in which the sample is located. Two examples value are, for example, "AA", meaning collectors issued by Angus Australia, and "ZOE", meaning collectors issued by Zoetis. For newly supplied samples (where "SAMPLE_TYPE" is "H", "T", etc) this field is optional. For existing samples (where "SAMPLE_TYPE" is "E"), this value must be left empty. For new hair samples, the allowed values are *"AA"* for Angus Australia, *"NAA"* for Neogen and *"ZOE"* for Zoetis. For TSUs, the onle value that is allowed is "TSU" for AllFlex TSU collectors. For tissue and semen samples, the "COLLECTOR_TYPE" field must be empty.

The fourth field is the *"ANIMAL_ID"* field. This field indicates the animal to which the sample belongs, and must be the Angus Australia animal ident. For both newly supplied and existing samples, this field is required. It is required for newly supplied samples so that we know the animal to which the sample belongs. It is required for existing samples for verification purposes. For example, sample barcodes are not necessarily unique – two samples with barcode "123456" could exist, but if so, they would usually belong to different animals.

The fifth field is the "STORE_ONLY" field. This field indicates whether the sample should be stored for future testing. A value of "X" means store for future testing. An empty value means do not store for future testing. Note that if a value of "X" is required, all variable fields (see below) for this row must be empty. Any other value is an error. It is also an error for an existing sample (where "SAMPLE_TYPE" is "E") to be "STORE_ONLY".

Variable Fields

The CSV file should also contain variable fields. These are in addition to the fixed fields above, and indicate the tests to be performed on the sample. In all cases, a value of "X" means perform the test while an empty value means do not perform the test. Any other value is an error.

The test fields uses a short mnemonic code to identify the test. These codes are prefixed with the laboratory code and separated by a dash, with "NAA" for Neogen and "ZOE" for Zoetis.

The remainder of the code is usually an acronym of the test name. For example, the genetic condition *Developmental Duplications* is often abbreviated "DD". The code for Neogen to test for *Developmental Duplications* is "NAA-DD" while the code for Zoetis is "ZOE-DD".

Note, however, that some codes are not always immediately obvious. For example, the code for a BVD re-test through Neogen is "NAA-BVDRE". A list of available test codes is available from the Angus Australia website.

Each order is assigned a single lab. It is an error to have tests for multiple labs within the same order.

Example

Below are two screenshots of a valid CSV file. The first is open in a spreadsheet program (*LibreOffice*). The second is open in a plain text editor (with line numbers to the left).

	A	B	C	D	E	F	G	H
1	HEADER	SAMPLE_TYPE	SAMPLE_BARCODE	ANIMAL_ID	STORE_ONLY	NAA-GS	NAA-DD	NAA-AM
2		H	AUAA-0000001	ABCX1	X			
3		H	AUAA-0000002	ABCX2		X	X	X
4		H	AUAA-0000003	ABCX3			X	X
5		H	AUAA-0000004	ABCX4			X	
6		S	AUAA-0000005	ABCX5		X	X	
7		E	123456	ABCX6			X	

```
HEADER,SAMPLE_TYPE,SAMPLE_BARCODE,ANIMAL_ID,STORE_ONLY,NAA-GS,NAA-DD,NAA-AM
,H,AUAA-0000001,ABCX1,X,,,
,H,AUAA-0000002,ABCX2,,X,X,X
,H,AUAA-0000003,ABCX3,,,X,X
,H,AUAA-0000004,ABCX4,,,X,
,S,AUAA-0000005,ABCX5,,X,X,
,E,123456,ABCX6,,,X,
```

The contents of the file is described below:

- Line 1 contains headers.

- Lines 2-7 contain data.

- A new hair sample is being provided for animals ABCX1, ABCX2, ABCX3 and ABCX4.

- A new semen sample is being provided for animal ABCX5.

- A sample barcode is being provided for all new samples. In this example, the sample barcodes are those provided by the Angus Society of Australia in sticker and/or collector form. While the Angus barcodes are preferred, any barcode is acceptable, for example, the barcodes pre-printed on Zoetis hair collectors, RFID-based barcodes for AllFlex TSU collectors, etc.

- An existing sample should be used for ABCX6, specifically, the sample with barcode 123456.

- The hair sample for ABCX1 should be stored for future use. No testing will be performed.

- Animals ABCX2 and ABCX5 should have an AngusGS produced.

- Animals ABCX2, ABCX3, ABCX4, ABCX5 and ABCX6 are all being tested for *Developmental Duplications*.

Animals ABCX2 and ABCX3 should be tested for *Anthrogryposis Multiplex.*

The Parent Verifications CSV File

Overview

The *Parent Verifications* CSV file contains a list of calves to parent verify, along with the sires and dams to include in each verification. This *Parent Verifications* CSV file is optional.

Required Fields

This CSV requires three fields, two of which can be duplicated. Note that unlike the above CSV file, there are no *variable* fields.

The first field is the "CALF_ID" field. This field indicates the calf that is to be parent verified.

The second field is the "SIRE_ID" field. This field indicates the sire against which the calf is to be parent verified. This field can be specified multiple times to test the calf against multiple sires.

The third field is the "DAM_ID" field. This field indicates the dam against which the calf is to be parent verified. Like the "SIRE_ID" field, this field can be specified multiple times to test the calf against multiple dams.

Note that for all fields, the animal (calf, sire or dam) must have either a SNP profile recorded in Angus Australia's database, or a request for a SNP profile in the previous CSV file. Note also that only SNP-based parent verifications are performed via this spreadsheet. You cannot use this CSV file to request parent verification using older MIP (microsatellite) technology.

Example

Below are two screenshots of a valid CSV file. The first is open in a spreadsheet program (*LibreOffice*). The second is open in a plain text editor (with line numbers to the left).

	A	B	C	D	E
1	HEADER	CALF_ID	SIRE_ID	SIRE_ID	DAM_ID
2		ABCX1	CBAZ1	USA987	CBAZ7
3		ABCX2	CBAZ1		

```
1    HEADER,CALF_ID,SIRE_ID,SIRE_ID,DAM_ID
2    ,ABCX1,CBAZ1,USA987,CBAZ7
3    ,ABCX2,CBAZ1,,
```

The contents of the file is described below:

- Line 1 contains the headers.

- Lines 2 and 3 contain parent verification requests.

- Line 2 is requesting the parent verification of calf ABCX1 against two sires, CBAZ1 and USA987, and one dam, CBAZ7.

- Line 3 is requesting the parent verification of ABCX2 against one sire, CBAZ1.

Additional Notes

If a CSV file is provided electronically to Angus Australia (for example, via email) and *not* through the angus.tech website, we also require a physical print out of the CSV file. This print out must contain the member's signature agreeing to Angus Australia's DNA terms and conditions, and must be posted along with the samples to Angus Australia. If samples arrive at Angus Australia for an electronically supplied CSV file, and no physical print out is provided, or is provided without a signature agreeing to the terms and conditions, the order will not be processed until such details are provided.

The member will be charged according to the DNA fee schedule, available on the Angus Australia website https://www.angusaustralia.com.au/. This includes any discounts that may be available, e.g., for ordering addon tests.

Samples and tests undergo a quality assurance process. During this process, poor quality samples may discarded and tests requests declined. For this reason, the spreadsheet represents *requests* for testing and does not guarantee that the test will be performed. Prices are calculated only according to those samples and tests that are approved.

A list of current test codes is available from the Angus Australia website.

If you have any other questions, please contact an Angus Australia staff member.