Requesting and receiving genomic evaluations
Tech Note 21

HIGHLIGHTS

- DataGene provides a routine genomic evaluation service.
- A Genomic Service Provider is a company that supplies genotypes and animal information and is authorised to receive results.
- To deliver a quality result, specific pieces of information are required from a Genomic Service Provider, in a timely manner.

Genomic Evaluation in Australia

DataGene delivers genomic evaluations for Holstein and Jersey males and females. Individual evaluations are conducted for a full range of traits, including:

- Indices (Balanced Performance Index, Health Weighted Index, Australian Selection Index)
- Yield
- Type
- Workability
- Daughter Fertility, Mastitis Resistance and Somatic Cell Count
- Survival (Longevity)
- Heat Tolerance
- Gestation Length
- Calving Ease
- Feed Saved

Genomic evaluations for Aussie Reds are expected to be released for limited traits from December 2020.

A parentage discovery service checks the sire and dam of each animal and will identify the correct parent (if known).
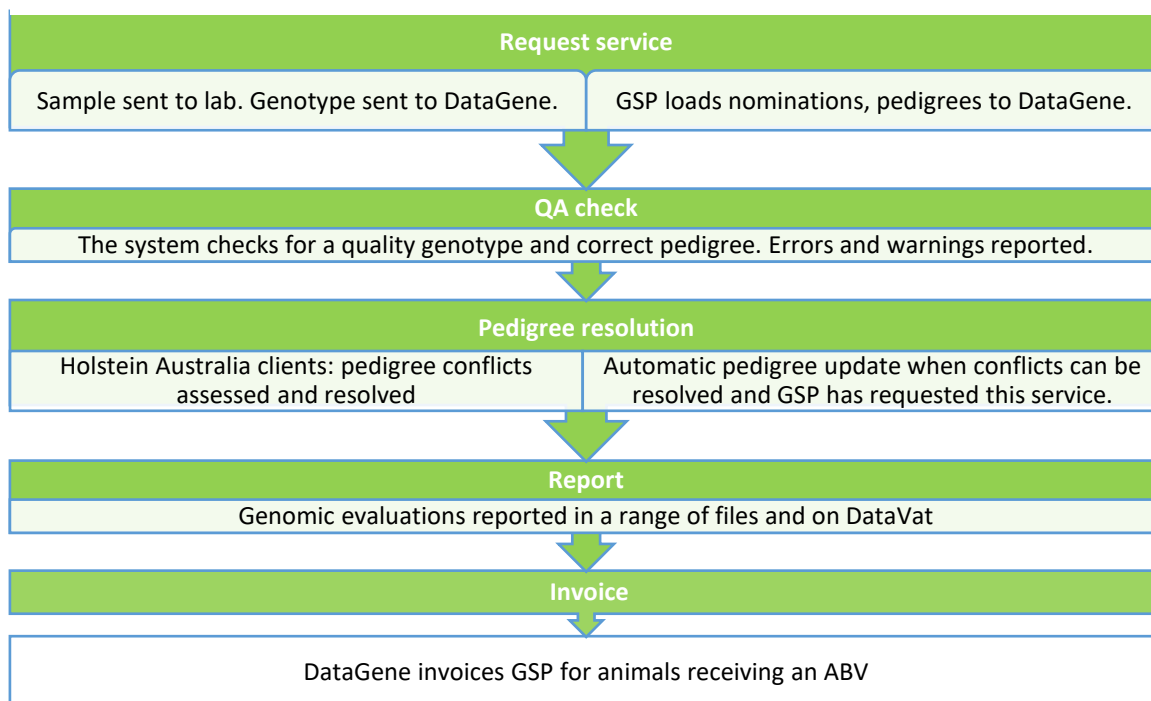
Haplotypes are called, including JH1, HH1, HH2, HH3.

Genomic service providers

A Genomic Service Provider is a company that supplies genotypes, nominations and animal information to DataGene and receives genomic evaluation results for their clients under a service agreement. Genomic Service Providers include specialist genomic labs, bull companies or breed associations. Services are available to companies and animals based in Australia or overseas. To request a service or receive results directly from DataGene, a company must be a Genomic Service Provider.

A summary of the steps involved in the genomic service are outlined in Figure 1.

Figure 1: Genomic process for Genomic Service Providers (GSP)

Request service	
Sample sent to lab. Genotype sent to DataGene.	GSP loads nominations, pedigrees to DataGene.

QA check
The system checks for a quality genotype and correct pedigree. Errors and warnings reported.

Pedigree resolution	
Holstein Australia clients: pedigree conflicts assessed and resolved	Automatic pedigree update when conflicts can be resolved and GSP has requested this service.

Report
Genomic evaluations reported in a range of files and on DataVat

Invoice
DataGene invoices GSP for animals receiving an ABV

Frequency of genomic evaluations

Genomic evaluations are conducted weekly, with a few exceptions due to holidays. Data (pedigree, genotypes, nominations) need to be loaded by Thursday 5 pm AEST to be included in the release 12 days later. Results are delivered every Tuesday with some exceptions due to public holidays.

DataGene and Dairy Australia have invested in new software that has improved service delivery and enables the increased frequency of evaluations.

Supplying information to DataGene

Genomic service providers and labs deliver data to DataGene via the user interface for the Central Data Repository (CDR). GSPs can look up the National ID of Australian cows using the cow lookup facility, which also creates National IDs for new cows. Overseas animals need to be loaded onto CDR with an International ID or herdbook ID and ancestry.

A nomination file links a sample ID to an animal and identifies the GSP and the party to be invoiced (generally the same). The cow lookup facility prepopulates a nomination file, to which sample IDs need to be added.

Receiving Information from DataGene

ABV results and genomic quality reports are delivered in two main ways:
1. Web portal known as DataVat
2. Delivered to your ftp or file share site

Holstein Australia's role

DataGene works closely with Holstein Australia to make it easier for clients to facilitate genomic evaluation. Farmers and organisations that are not GSPs should discuss their needs with Holstein Australia.

Holstein Australia services include

- processing samples and sending to the lab for genotyping
- gathering pedigree data
- assessing parentage
- preparing and loading pedigree and nomination files onto CDR
- preparing genotypes received from overseas for inclusion in genomic evaluations
- returning results

More details: Holstein Australia
genomics@holstein.com.au Ph: 03 9835 7600

Information required

Conducting the genetic evaluation requires three pieces of information; pedigree, nomination and genotype. The most common cause of 'no result' is a missing piece of information.

1. Pedigree record – Animal's own national ID plus the Sire and Dam's national ID.
- Herd recorded animals with complete pedigrees – in this case, the National ID of the animal is all that is required as the pedigree is stored in the Centralised Data Repository (CDR). You can find national IDs from herd test centres, many on-farm software programs, farmer computer records and breed associations. The cow lookup facility can be used to find National IDs or create new ones for as yet unrecorded females. Holstein Australia can also create National IDs.
- Non herd recorded animals – where a herd has not been herd recording, a National ID can be generated with the cow lookup facility or by Holstein Australia.
- AI bulls - In the case of a candidate AI bull, three generations of pedigree are required, NASIS clients can enter bulls directly into the CDR.
- Other bulls – GSPs and NASIS clients can create new bulls in the CDR.

Genomic breeding values can be calculated for animals without parents, but the reliability of the breeding values increases with the addition of ancestors.

Pedigree records are submitted to the CDR using one of these formats
a) DIF 102/105 files (for domestic animals)
b) Interbull 200 file (for overseas animals)
c) cow excel
d) bull excel
e) DIF115 for overseas cows
Sample formats are shown below and available at https://datagene.com.au/node/1451.

2. Genotype
The Genotype is generally sent directly from a lab to DataGene. Genotypes that are passed on from overseas should be submitted to a GSP who can load them onto the CDR.

Genotypes are accepted as ggg/gms/gss in TOP TOP, US SNP order files. Raw genotypes must include TOP TOP.

3. Nomination or Linking
A nomination file is used to link the genotype file to the pedigree of the animal.

The new format nomination file has only four columns; the animal National ID, the sample ID, the GSP and the party to be invoiced.
Sample formats are shown below and available at https://datagene.com.au/node/1451.

Reasons for failed results or warning
A warning is reported, under secure access to www.datavat.com.au if the animal has a missing parent. An animal may fail to get a genomic breeding value for several reasons

1. No genotype could be linked to the animal Genotype may not have been submitted by the lab or the nomination file may not correctly identify the link
2. Genotype failed QA
3. Genotype is inconsistent with a parent's genotype
4. Genotype is for a female, but animal is male
5. Genotype does not have enough markers for genomic evaluations (it can still be used for parentage)
6. No request was received to calculate genomic breeding values for the animal
7. Genomic breeding values are not calculated for the animal's or its sire's breed
8. The genotype is based on a chip that has not been accepted by DataGene

Figure 2: Format for the Interbull 200 file that is used to provide pedigree (usually of non Australian animal) ref http://www.interbull.org/ib/file200
For full details, including notes, refer to Data Interchange Formats described at datagene.com.au

Field No.	Field Name	Start Column	Length	Numeric /Alpha	Comments
1	Record Type	1	3	N	Value = 200
2	International ID of the animal	5	19	A	see note below
3	International ID of Sire	25	19	A	same as animals IID see note below
4	International ID of dam	45	19	A	same as animals IID see note below
5	Date of Birth	65	8	N	yyyymmdd
6	Status of animal	74	2	N	see note below
7	Date of birth of first AI daughter	77	8	N	yyyymmdd
8	Name	86	30	A	
9	International ID of the animal	117	19	A	see note below
10	Country Code	137	3	A	See note 2
11	Bull owner code	141	3	A	See note 6

RECORD LENGTH =143 bytes

Text file format.

Status of bull: 00 unknown; 10 bull randomly sampled through an official AI scheme; 20 other bull. Records with "20" in this field will be excluded from the international evaluation, unless type of proof is "21".

International ID
The International ID as designated by Interbull has the following format
Breed	3 characters (eg, HOL, JER, AYS, GUE)
Country	3 characters (eg, AUS, USA, CAN - see Note 2 for a full list of codes)
Sex	1 character (M or F)
Within-Country ID	12 characters (right justified, zero filled)

Figure 3: File format for the DIF 102 file that provides pedigree data for a cow
(usually an Australian cow). Ref and notes available at https://datagene.com.au/DataInterchangeFormats

Field No.	Field Name	Start Column	Length	Numeric /Alpha	Comments
1	Record Type	1	3	N	Value = 102
2	Record Version Number	4	1	A	Value = 1
	Herd ID				
3	National Herd ID	5	7	A	See Note 8
	Cow Identity				
4	National Cow ID	12	9	A	See Note 3
5	Within-Herd Cow ID	21	6	N	
	Herdbook ID				
6	Country Code	27	3	A	See Note 2
7	Herdbook Number	30	12	A	See Note 2
8	NLIS Animal ID	42	16	A	
9	NLIS Tag Radio Frequency	58	16	A	
10	Breed	74	4	A	See Note 1
11	Birth date	78	8	N	yyyymmdd
	Pedigree details				
12	Sire National ID	86	9	A	See Note 3
13	Dam National ID	95	9	A	See Note 3
14	MGS National ID	104	9	A	See Note 3 (Required by DataGene if Dam ID is unavailable, and MGS is available)
	Transfer Details				See explanation below
15	Transfer-in date	113	8	N	yyyymmdd
16	National ID of Herd Transferred from	121	7	A	See Note 8
	Cow Name				
17	Long	128	40	A	
18	Short	168	16	A	
	Cow status codes				
19	Animal termination code	184	2	A	Sold and dead codes - see Note 5
20	Animal termination date	186	8	N	yyyymmdd
21	Sire verification flag	194	1	A	Value Y = yes; N = no

RECORD LENGTH = 194 bytes

Transfer of Cows between Recorded Herds
DataGene needs to have the capacity to analyse lactations with the herd in which the lactation occurs. If a cow is transferred from one herd to another, the details required are the date of transfer of a cow into a herd and the National Herd ID of her previous herd. The vast majority of cows are never transferred, and for these cows the two fields should be left blank.

Essential fields for DataGene are 1, 2, 3, 4, 7, (6 if 5 is non-blank), (21 if 12 is non-blank). Fields 11 and 12 are also required for a cow to receive an ABV. Other fields are strongly recommended. Assumed sort order with all fields in ascending order : Fields 1, 2, 3, 4.

Figure 4: File format for the DIF 105 file that provides pedigree data for a bull

Field No.	Field Name	Start Column	Length	Numeric /Alpha	Comments
1	Record Type	1	3	N	Value = 105
2	Record Version Number	4	1	A	Value = 3
	Bull Identity				
3	Bull National ID	5	9	A	See Note 3
	Bull Herdbook ID				
4	Country Code	14	3	A	See Note 2
5	Herdbook Number	17	12	A	See Note 2
6	Local Bull ID	29	15	A	
7	Date of Birth	44	8	N	yyyymmdd
8	Bull Breed	52	4	A	See Note 1
	Pedigree Details				
9	Sire National ID	56	9	A	See Note 3
10	Dam National ID	65	9	A	See Note 3
11	MGS National ID	74	9	A	See Note 3
12	Bull name	83	40	A	
	NASIS Bull Details				
13	NASIS Primary ID	123	7	A	
14	Bull ID	130	12	A	
15	Bull Owner Code	142	3	A	See Note 6
16	International ID	145	19	A	Interbull format - see note below
17	PT Sampling Code	164	1	A	
18	Date First Semen Available	165	8	N	yyyymmdd
19	Genetic Codes	173	8x3	A	Up to 8 three-character codes - see note 10
20	NASIS Active Sire Code	197	1	A	A = active, R = restricted, W = warning of a possible conflict with the ID of another bull, blank = not active
21	Common name 1	198	12	A	Name used in marketing of bull
22	Common name 2	210	12	A	Name used in marketing of bull
23	Date Sexed Semen Available	222	8	N	yyyymmdd(blank=no sexed semen available)

RECORD LENGTH = 229 bytes

International ID
The International ID as designated by Interbull has the following format

Breed	3 characters (eg, HOL, JER, AYS, GUE)	
Country	3 characters (eg, AUS, USA, CAN - see Note 2 for a full list of codes)	
Sex	1 character (M or F)	
Within-Country ID	12 characters (right justified, zero filled)	

Essential fields for DataGene are 1, 2, 3, 8. Other fields are strongly recommended.
Assumed sort order with all fields in ascending order: Fields 1, 2, 3.

Dairy Australia
Your Levy at Work

DataGene
Solutions for Herd Development

Figure 5: File format for the DIF 115 file that provides international cow pedigree record

Ref and notes available at https://datagene.com.au/DataInterchangeFormats

Field No.	Field Name	Start Column	Length	Numeric /Alpha	Comments
1	Record Type	1	3	N	Value = 115
2	Record Version Number	4	1	A	Value = 1
	Animal Identity				
3	National ID	5	9	A	See Note 3
	Herdbook ID				
4	Country Code	14	3	A	See Note 2
5	Herdbook Number	17	12	A	See Note 2
6	International ID	29	19	A	Interbull format - see note below
7	Date of Birth	48	8	N	yyyymmdd
8	Breed	56	4	A	See Note 1
9	Name	60	40	A	
	Pedigree Details				
10	Sire National ID	100	9	A	See Note 3
11	Dam National ID	109	9	A	See Note 3
12	MGS National ID	118	9	A	See Note 3

RECORD LENGTH = 126 bytes

This record exists to capture details on foreign cows for inclusion in the pedigree of bulls and cows used in Australia.

International ID

The International ID as designated by Interbull has the following format

Breed	3 characters (eg, HOL, JER, AYS, GUE)
Country	3 characters (eg, AUS, USA, CAN - see Note 2 for a full list of codes)
Sex	1 character (M or F)
Within-Country ID	12 characters (right justified, zero filled)

Essential fields are 1, 2, 3, 6. All other fields are strongly recommended.
Assumed sort order with all fields in ascending order : Fields 1, 2, 3.

Dairy Australia
Your Levy at Work

DataGene
Solutions for Herd Development

Figure 6: File format to upload bulls in a batch

Ref and notes https://datagene.com.au/DataInterchangeFormats
Sample file available at https://datagene.com.au/node/1451

Figure 7: File format to upload cows in a batch

Ref and notes https://datagene.com.au/DataInterchangeFormats
Sample file available at https://datagene.com.au/node/1451

Figure 8: Genomic nomination – new format

Ref and notes https://datagene.com.au/DataInterchangeFormats
Sample file available at https://datagene.com.au/node/1451

New format nomination

Required Fields	Char	Comments
National ID	9	See Note 3
Sample ID		
Reciepeint code	3	Must be 3 characters
Bull Owner code	3	Must be 3 characters

Text file format. Fields need to be separated from a tab (/t).

Example;

Figure 9: Look up cows

Input format for cow lookup

Required Fields	Char	Comments
National Herd ID	7	Must exist in CDR
Within-herd Cow Number		Integer between 1 and 999999
Breed Code	4	Must be 4 characters
Birth Date YYYYMMDD		Valid date, cannot be a future date
Name	40	Could be same as within-herd cow No.

Optional Fields		
Sire National ID	9	
Dam National ID	9	
MSG National ID	9	

Comma delimited or csv file format.

Field Sourced from party
Source party: from the GSP user logged in

A match is made if, all three rules below are TRUE:
a. Within-herd Cow Number matches a cow in the herd specified.
b. The matching animal is female.
c. If a birth date is recorded in CDR it must be within 2 years.

Note:
a. A match with a male is impossible as a male does not have a herd and within-herd ID.
b. There is no match on breed code and name.
c. The DOB must be within 2 years of the DOB on CDR. If CDR has no DOB, then a match is assumed.
d. If a match is found, the data on the database is NOT updated with information from the input file. If an update is required, use the cow management UI or load a DIF102 or cow excel record.

Figure 10: DIF565 Parentage Discovery Report

565 output from Parentage Discovery Service

Field No.	Field Name	Start Column	Length	Numeric /Alpha	Comments
1	Record Type	1	3	N	Value = 565
2	Record Version Number	4	1	A	Value = 1
3	Report Type	5	1	A	Q=QA, P=Parents discovered
	Animal IDs				
4	Animal National ID	6	9	A	See DIF note 3
5	Laboratory Code	15	2	A	See note below
6	Animal Sample ID	17	16	A	e.g. sample bar code
	Sample QA and Consistency				
7	Issue Code 1	33	2	N	See note below
8	Issue Code 2	35	2	N	See note below
9	Issue Code 3	37	2	N	See note below
10	Issue Code 4	39	2	N	See note below
11	Issue Code 5	41	2	N	See note below
12	Issue Code 6	43	2	N	See note below
	Parentage Report Details				
13	Initial Report, Update or Revision	45	1	A	I=Initial; U=Update; R=Revision
14	SNP Density used to Discover Sire	46	2		See note below
15	SNP Density used to Discover Dam	48	2	A	See note below
16	Number of Previous Samples for Animal	50	2	N	
17	Discovery Herd National ID	52	7	A	See DIF note 8
18	Number of Sires Discovered	59	2	N	
19	Number of Dam Discovered	61	2	N	
	Parents Discovered				
20	Laboratory Code for Sire Sample	63	2	A	See note below
21	Sire Sample ID	65	16	A	
22	Sire National ID	81	9	A	
23	Sire Local Animal Name	90	15	A	Local bull ID
24	Sire National Herd ID	105	7	A	
25	Sire Birth Date	112	8	N	yyyymmdd
26	Laboratory Code for Dam Sample	120	2	A	See note below
27	Dam Sample ID	122	16	A	
28	Dam National ID	138	9	A	
29	Dam Local Animal Name	147	15	A	Within-Herd Cow ID
30	Dam National Herd ID	162	7	A	
31	Dam Birth Date	169	8	N	yyyymmdd
	Additional Possibility for Parents				
32	Laboratory Code for Sire Sample	177	2	A	See note below
33	Sire Sample ID	179	16	A	
34	Sire National ID	195	9	A	
35	Sire Local Animal Name	204	15	A	Local bull ID
36	Sire National Herd ID	219	7	A	
37	Sire Birth Date	226	8	N	Yyyymmdd
38	Laboratory Code for Dam Sample	234	2	A	See note below
39	Dam Sample ID	236	16	A	
40	Dam National ID	252	9	A	
41	Dam Local Animal Name	261	15	A	Within-Herd Cow ID
42	Dam National Herd ID	276	7	A	
43	Dam Birth Date	283	8	N	Yyyymmdd
44	Comment	291	255	A	

Record length = 545 bytes

Dairy Australia
Your Levy at Work

DataGene
Solutions for Herd Development

Figure 11: DIF561 Parentage Discovery Sample Description

561 nomination for Parentage Discovery Service only

| Field No.	Field Name	Start Column	Length	Numeric /Alpha	Comments
1	Record Type	1	3	N	Value = 561
2	Record Version Number	4	1	A	Value = 1
3	Laboratory Code	5	2	A	See note below
4	Electronic sample ID	7	16	A	e.g. sample bar code
5	Local animal name	23	15	A	Local bull ID or within-herd cow ID
6	National Animal ID	38	9	A	See DIF note 3
7	Sex of animal	47	1	A	Value M = Male; F = Female
8	Date of Birth	48	8	N	yyyymmdd
9	Date of Birth Status Flag	56	1	N	See note below
10	National Herd ID	57	7	A	See DIF note 8
11	Lab DNA Analysis Requested	64	2	N	See note below
12	Parentage Discovery Required	66	1	A	Value Y = Yes, N = No
13	Service Provider Code	67	3	A	See DIF note 6
14	Physical sample	70	1	A	See note below
15	Sample type	71	1	A	See note below
16	Third party billing code	72	3	A	See note below

Record length = 74 bytes

More information

Contact the genotyping service desk
E: genotypes@datagene.com.au
www.datagene.com.au.   October 2020
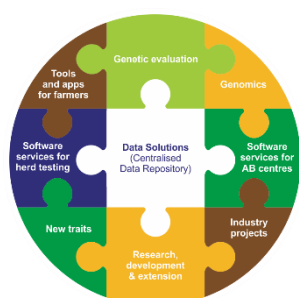
About DataGene

DataGene is an independent and industry-owned organisation responsible for driving genetic gain and herd improvement in the Australian dairy industry. DataGene performs pre-competitive herd improvement functions such as genetic evaluation, herd testing and herd improvement software development, data systems and herd test standards. DataGene is a Dairy Australia and industry collaboration. www.datagene.com.au Ph 1800 841 848 E: abv@datagene.com.au