Process for requesting and receiving genomic evaluations

HIGHLIGHTS

- DataGene provides a monthly genomic evaluation service. Service schedules are expected to change in line with new software.
- A Genomic Service Provider is a company that supplies genotypes and animal information and is authorised to receive results.
- There are specific pieces of information required from a Genomic Service Provider, in a timely manner, to deliver a quality result.

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, Type Weighted Index, Australian Selection Index)
- Yield traits
- Type traits
- Workability traits
- Daughter Fertility and Somatic Cell Count
- Survival (Longevity) and Residual Survival
- Heat Tolerance
- Gestation Length (coming soon)
- Calving Ease (coming soon)

A parentage discovery service checks the sire and dam of each animal and will identify the correct parent (if known).

A limited range of Haplotypes are called, including JH1, JH2, HH1, HH2, HH3.

Who is a Genomic Service Provider?

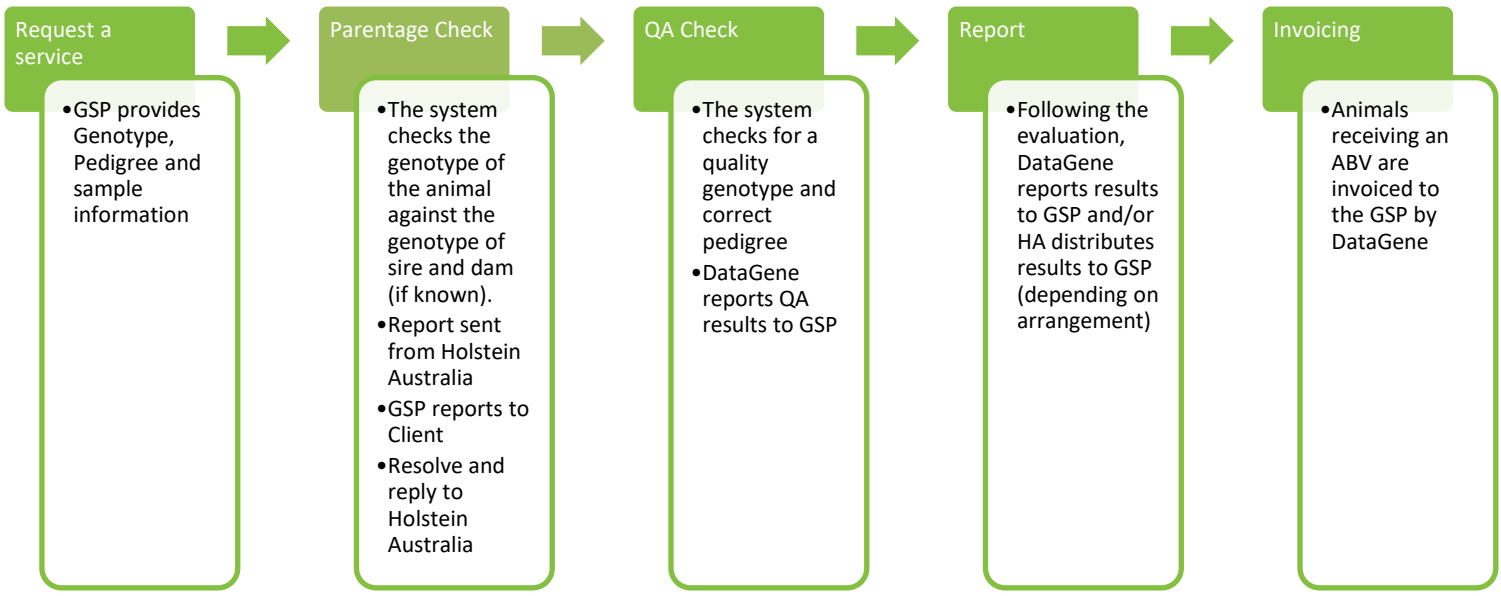
A Genomic Service Provider is a company that supplies genotypes and animal information to receive genomic evaluation results for their clients. Generally, Genomic Service Providers are labs, bull companies or breed associations. Companies can be based in Australia or overseas. The animals that are tested can be based in Australia or overseas.

To request a service or receive results directly from DataGene, a company must be a Genomic Service Provider.

Genomic Service Step

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

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



DataGene
Solutions for Herd Development

Frequency of Genomic Evaluations

Currently, genomic evaluations are conducted monthly. Data (pedigree, genotypes, nominations) needed to be loaded before the third Monday of each month. Results are delivered the first Tuesday of the month. There are some exceptions due to public holidays.

DataGene and Dairy Australia have invested in new software that will help improve service delivery and increase the frequency of evaluations. The new software is close to completion and it is expected that service delivery times will improve.

Supplying Information to DataGene

DataGene receives information in three main ways:

1. Email to genotypes@datagene.com.au
2. Ftp site
3. Proprietary on-line large file services such as DropBox, WeTransfer,

Files can be accepted by Excel Spreadsheet, DIF (Data Interchange Format) or Interbull 200 flat text format.

Receiving Information from DataGene

ABV results are delivered in three main ways:

1. Web portal
2. Delivered to your ftp or file share site
3. Emailed to the designated email address.

The role of Holstein Australia

DataGene works closely with Holstein Australia to make it easier for clients to facilitate genomic evaluation. Clients who do not wish to load pedigree and nominations using DataGene's web platform, on their own, 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 in advance of the genetic evaluation process
- preparing and loading pedigree and nomination files
- returning results

Contact Holstein Australia for more details.

What Information is Required?

Conducting the genetic evaluation requires three pieces of information; pedigree, sample information 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, pedigree is straight forward as the National ID of the animal is all that is required as the pedigree is stored in the Centralised Data Repository. National ID's for animals are generated by herd test centres or Holstein Australia. You can obtain national IDs from herd test centres, many on-farm software programs, farmer computer records or breed associations.
- Non herd recorded animals - Where a herd has not been herd recording, a National ID can be generated by Holstein Australia.
- AI bulls - In the case of a candidate AI bull, three generations of pedigree are required.

Pedigree records are submitted to DataGene using one of two formats
a) DIF 102/105 files (for domestic animals)
b) Interbull 200 file (for overseas animals)
Other formats may be considered but attract higher data processing fees.

2. Genotype

The Genotype is generally sent directly from a lab to DataGene.

3. Nomination or Linking

In order to connect the genotype file to the pedigree of the animal, a nomination file is used to link the pieces of information.

Nominations are submitted to DataGene using the DIF 481 file.
Other formats may be considered but attract higher data processing fees.

Figure 2: File format for the Interbull 200 file that is used to provide pedigree (usually of non Australian animal)

ref http://www.interbull.org/ib/file200

File Format 200 for Dairy Pedigree

Starting position	Field	Format	Example
1	Record type	Character 3	200
	International ID of ANIMAL		
5	Breed of animal	character 3	HOL
8	Country of first registration of animal[1]	character 3	CAN
11	Sex	character 1	M
12	ID number of animal[2]	character 12	000000A12345
	International ID of Sire of ANIMAL		
25	Breed of sire of animal	character 3	HOL
28	Country of first registration of sire of animal[1]	character 3	CAN
31	Sex	character 1	M
32	ID number of sire of animal[2]	character 12	556912367589
	International ID of Dam of ANIMAL		
45	Breed of dam of animal	character 3	HOL
48	Country of first registration of dam of animal[1]	character 3	CAN
51	Sex	character 1	F
52	ID number of dam of animal[2]	character 12	123569874521
65	Birth date of animal[4] (YYYYMMDD)	integer 8	19870215
74	Status of animal[3]	Integer 2	10
77	Birth date of first AI daughters (YYYYMMDD)	Integer 8	19890314
86	Name of animal	character 30	Cantarello
	National ID of ANIMAL		
117	Breed of animal	character 3	HOL
120	Country	character 3	CAN
123	Sex	character 1	M
124	ID number of animal[2]	character 12	000000A12345
137	Country sending information	character 3	CAN
Character encoding: UTF-8			

1. Identification in the country of first registration, as known in the country sending this information.

2. All ID numbers: Registration numbers, right justified, leading blanks as zeros.

3. Status of bull: **00** unknown; **10** bull randomly sampled through an official AI scheme; **20** other bull. Records

4. Check the animal's birth date

- Has to be reported in the format YYYYMMDD
- If you know only the year of birth then enter it as YYYY0000
- If you know year and month of birth then enter them as YYYYMM00
- Missing birth dates are coded as 00000000 (or blanks or a single 0)

Figure 3: File format for the DIF 102 file that provides pedigree data for a cow (usually an Australian cow).

Ref 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

Ref https://datagene.com.au/DataInterchangeFormats

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.

Figure 5: File format for the DIF 481 file that links a sample to an animal.

Ref https://datagene.com.au/DataInterchangeFormats

Data Format 481 V3 **Genotype Nominations File**

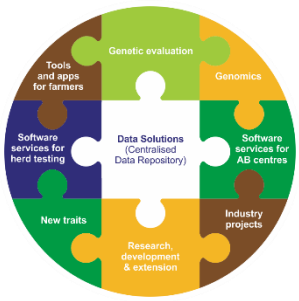
Field No.	Field Name	Start	Length	Numeric /Alpha	Comments
1	Record Type	1	3	N	Value = 481
2	Record Version Number	4	1	A	Value = 3
	Sample details				
3	Laboratory code	5	2	A	See note below
4	Lab DNA analysis requested	7	2	N	See note below
5	Electronic sample ID	9	16	A	e.g. sample bar code
6	Sample type	25	1	A	Value B = blood, E = ear plug, H = hair, S = semen (males only)
7	Date of sample	26	8	N	yyyymmdd
	Animal Details				
8	NLIS Tag Radio Frequency	34	16	A	
9	National ID	50	9	A	See DIF note 3
10	Local animal name	59	15	A	Local bull ID or within-herd cow ID
11	Sex of animal	74	1	A	Value M = male, F = female
12	National Herd ID	75	7	A	This field is essential for local females
13	DPC Code	82	1	A	Data Processing Centre - see DIF note 4
	Supplier and recipient details				
14	Genotyping service provider code	83	3	A	See DIF note 6
15	Designated recipient of results	86	70	A	Genotyping service provider code or recipient email address
16	Genotyping requested by	156	3	A	See DIF note 6

RECORD LENGTH = 158 bytes

Essential fields for all animals are 1, 2, 3, 4, 5, 9, 10, 11, 14, 15 and 16. Fields 12 and 13 are essential for females. Other fields should be provided when available.

More information

Michelle Axford
Ph 0427573330
E: maxford@datagene.com.au
www.datagene.com.au.
December 2018



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 03 9032 7191 E: abv@datagene.com.au