DIGAD Data Standard for Submitting Parentage and Breed Recording and Genetic Testing Data to the Dairy Industry Good Animal Database

('Genetic testing standard')

Contents

Log of Amendments

Standard issued by NZAEL Board the 7 day of May 2024, Version 30 April 2024

Amendment	Effective date	Version	Method of Authorisation

Summary of amendments[1]:

[1] The type of change – substantive or non-substantive is provided with a description of the details for each approved amendment.

1 Introduction (Informative)

Breeding genetically superior animals leads to genetic gain. Accurate pedigree recording for an animal and its ancestors is important so animals are ranked correctly. This accuracy is more important in a multibreed context:

- where the recorded pedigree of an animal is used to estimate its breed proportions, and
- where accurate phenotypic measures of animals with precisely recorded breed proportions are used to estimate breed effects.

Achieving world-leading rates of genetic gain, especially in a multibreed context, relies on accurately recording animals' parentage and therefore breed. Genetic testing using low-cost SNP chip technology enables SNP-based verification of parentage. In situations where an animal's parentage is unknown, genomic data on the offspring and their putative parents can be used to identify their parents.

Genomic selection has reduced artificial insemination breeding companies' reliance on costly and time-consuming progeny testing (see Appendix 1) to identify bulls with high genetic merit.

It has also significantly advanced the rate of genetic gain as it enables dairy cattle breeding programs to identify genetically superior animals at a much earlier age. The rate of genetic progress for economically important dairy traits can be approximately doubled by:

- increasing the accuracy of predicted genetic merit for young animals,
- shortening the generation interval due to the wider use of young, genetically superior males and females, and
- increasing the intensity of selection because breeders can use genomic testing to screen a larger group of candidate animals.

This standard covers the activities related to parentage and breed recording, breed proportions, and genetic testing. This is relevant to:

- delivering world-class genetic gain within the New Zealand dairy industry, and
- achieving high quality and transparent animal evaluations.

For more information on genetic gain and estimating breeding values in a multibreed context see Appendices 1 and 2.

2 Scope

This standard is a DIGAD Data Standard. It defines the requirements for data and associated metadata, and the processes and procedures for submitting these data to the Dairy Industry Good Animal Database (DIGAD) for use in animal evaluation (AE). It covers:

- recording parentage and breed; and
- genetic testing for:
    - parentage verification and discovery; and
    - genomic profiles for inclusion in AE (including genomic selection), for animals enrolled in the DIGAD.

3 Normative References

The following documents are referred to in the text in such a way that some, or all, of their content constitutes requirements of this standard. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

- ISO/IEC 17025 is an international standard that enables laboratories to demonstrate that they operate competently and generate valid results. Accreditation to the standard facilitates cooperation between laboratories and other bodies internationally by generating wider acceptance of results.
- DIGAD Data Standard Terminology which provides definitions of terms used in this standard.
- NAIT Standard, Animal Identification Devices (December 2020 - with amendments or replacements as gazetted by the Government of New Zealand). This standard is for the radio frequency devices attached to cattle and deer for NAIT tracing in New Zealand.

- ISO 11784:1996 Radio frequency identification of animals. This standard specifies the code structure used in radio frequency devices for identifying animals, and includes amendments published in 2004 and 2010.

4 Terms and Definitions

For the purposes of this standard, the following terms are of particular significance to this document[2]:

Animal durable code	The animal durable key with the suffix 'DK' left padded with 0's to make 12 digits, e.g. DK000012345678.
Animal durable key	The primary unique animal identification assigned to each animal when it is enrolled with DIGAD. The animal durable key is a lifetime identification that persists regardless of change to other mandatory or optional identifiers.
Animal evaluation (AE)	The process of fitting a model which describes the systematic and non-systematic factors and effects that influence observed performance. The model is fitted to performance records such as those from herd testing. The outcomes from AE include estimates of the effects and the likely nature of prediction errors. From a selection viewpoint, it is the estimates of the breeding values, and their associated reliabilities which are a function of prediction errors, that are of paramount importance. In addition to performance records such as from herd testing, modern animal evaluation typically makes use of pedigree information, genomic information and prior information from other sources. In its widest context, animal evaluation includes the process of combining across all traits in the national breeding objective, the estimates of breeding values with their economic values to produce an overall estimate of Breeding Worth (BW).
Animal identifiers	The industry identifiers used to identify and animal. In the NZS8100 Standard they are termed 'unique animal identifiers.
Approved fields of activity	The data fields that the Certified Data Provider has been certified to access and supply to DIGAD
Certification Body	Certification Body. The auditing organisation appointed by the NZAEL Board that audits and certifies individuals and organisations as Herd Record Providers and Certified Data Providers to NZAEL.
Certified Data Provider (CDP)	Individuals or organisations that are certified by the approved Certification Body as meeting the appropriate standard for the supply of non-regulated data or information to NZAEL. Each farmer (i.e. participant) can contract multiple Certified Data Providers.
Certified herd tester (CHT)	A person who is certified as a herd tester under Schedule 1 of the Dairy Industry (Herd Testing and New Zealand Dairy Core Database) Regulations. A CHT cannot be certified by the CDP Certification Body appointed under this standard. Each farmer (i.e. participant) can only contract one Certified Herd Tester.
Core-data	The data in or required by the Herd Testing Regulations to be contributed to the Core Database and any additional data pursuant to any amendment to the Act or the Herd Testing Regulations
Core database	As per the Dairy Industry Restructuring Act 2001, means the core database that comprises (a) Information provided to the manager of the core database under the Herd Testing Regulations 1958 or under the terms and conditions of any licence issued under those Regulations; and

[2] *Note: The terms and definitions will be in the terminology and the term in this section linked to that definition when published*

	(b) Information provided to the manager of the core database under any Regulations made under this Act.
DairyNZ Limited (DairyNZ)	The industry organisation that is funded by a dairy farmer levy to represent all New Zealand dairy farmers.
Dairy Industry Good Animal Database (DIGAD)	A database containing ancestry and performance data for every recorded dairy cow in New Zealand. This database is intended to contain all the data required for AE. DIGAD contains the core database and other data fields.
DIGAD CDP Interface Specification for Herd Record Providers	This document describes the electronic interface by which Herd Record Providers may routinely register dairy farm data with DIGAD.
DIGAD Data Providers	Individuals and organisations that submit data to DIGAD, including CHTs, herd record providers (HRPs), and certified data providers (CDPs).
DIGAD Data standards	The standards developed under the oversight of the DIGAD Data standards Committee. The standards set out the requirements for DIGAD Data Providers to be certified by the Certification Body.
Herd Test Standard	The New Zealand Standard NZS 8100 Dairy Herd Testing, which sets out the operational requirements for Certified Herd Testers and which is cited in the Dairy Industry (Herd Testing and New Zealand Dairy Core Database) Regulations 2001.
Herd Record Provider (HRP)	A DIGAD Data Provider certified by the approved Certification Body as meeting the appropriate standard for the supply of data or information to NZAEL. An HRP is contracted by a farmer to enrol animals, manage animal movements, and animal termination in DIGAD on behalf of the farmer. Each farmer (i.e. participant) can only contact one Herd Record Provider.
New Zealand Animal Evaluation Ltd (NZAEL)	The organisation that develops and delivers the technologies that publish the independent evaluation of all dairy animals using the Breeding Worth (BW) index.
Non-core data	The data supplied to DIGAD not defined as core-data and agreed to be provided to DIGAD under a data supply agreement with the NZAEL database manager.
Regulated fields	Core data fields that are required to be collected and supplied under the Dairy Industry (Herd Testing and New Zealand Dairy Core Database) Regulations 2001, or its successor, and in accordance with current version of the NZS8100: Dairy Herd Testing Standard.

5 Interpretation

For the purposes of this standard, the word ‘shall’ refers to requirements that are essential for adherence with the standard, while the word ‘should’ refers to practices that are advised or recommended.

The standard is considered normative (required) unless a section has been specifically identified as informative. Some sections include both informative and normative information. If a subsection within a section is not identified as informative then subsection is normative. Appendices are informative and provide additional information.

6 Structure of the Standard (informative)

The standard comprises parentage and breed recording in DIGAD, SNP based parentage analysis, requirements of a genotyping laboratory including sample submission, delivery of genomic data to the genomic database associated with DIGAD and its quality control when used in AE activities.

This standard provides an overview of the data requirements for submitting parentage and breed recording data to DIGAD whether via the Herd Test Standard or this DIGAD data standard.

To ensure consistency between the two standards, Standards NZ has permitted the use of the same terminology in this standard as in the Herd Test Standard.

7 Submission of Data

Data requirements for submitting parentage and breed recording data to DIGAD are detailed in one of two standards:

1. NZS 8100:2015, New Zealand Standard, Dairy herd testing, or its successor ('Herd Test Standard') which sets out the requirements and procedures for the collection and supply of core-data (i.e. regulated fields) to the core database.
2. The current standard, which sets out the requirements for processing data and procedures for the collection and supply of non-core data and provides an overview of the combined requirements.

A Herd Record Provider (HRP) typically administers a herd management system and manages and collates data for their client but is not bound by the regulations detailed in the Herd Test Standard.

A Certified Herd Tester (CHT) when providing regulated herd testing services to clients is required by law to ensure all regulated fields are collected and submitted to DIGAD's Core Database. Many of these fields are included as part of animal enrolment requirements for HRPs.

Consequently, there is an inextricable connection between HRPs and CHTs. Typically, a CHT will have a contractual agreement with an HRP to ensure regulated data are submitted on the CHT's behalf, so the CHT does not breach the regulations.

8 Parentage or Breed Recording (informative)

This standard sets the requirements to ensure that the parentage of each animal, and therefore its expected ancestry breed, is recorded correctly. Identifying the correct parentage of an animal is an important step for AE.

Parentage can be established by using:

- mating data,
- embryo data,
- calving records, and
- DNA testing and subsequent analyses.

Details for recording mating and embryo data, and calving records are available in the Herd Test Standard or in the 'DIGAD CDP Interface Specification for Herd Record Providers' ('DIGAD Interface Specifications') document.

This section focuses on parentage and the use of DNA testing to verify or discover parentage.

The breed composition of an animal can be determined if its parents are known and their ancestral founders have a known breed composition. The breed composition for animals with no known parents relies on the breed 16ths recorded for them (see following section).

The primary objective is that animals who are grossly different in performance due to breed are recorded as accurately as possible.

Parentage informs breed, so while the utmost focus should be on accurate recording of parentage, breed 16ths should continue to be recorded.

8.1 Recording of breed 16ths (informative)

An HRP provides putative breed 16ths (and genetic test) information when animals are enrolled as it is a contractual requirement between the manager of DIGAD and an HRP. An HRP normally enrols animals on behalf of the CHT and fulfills the CHTs legal obligations for providing the core data relating to breed 16ths.

The Herd Test Standard sets out the data requirements for this field. It does not require the breed 16ths to sum to one, but they cannot exceed one. Breed 16th data are only used for DIGAD and to inform AE when required.

Only an HRP, who is contracted to provide 'Animal Identification' data can submit this information through the 'Animal Event' interface.

Breed 16ths is recorded when an HRP creates an animal record or when a birth is recorded. The relevant breed is either derived from an animal's known parentage or where a farmer corrects records. Up to eight breeds may be recorded to define the breed fractions of an animal.

An HRP only updates a record after it is established in the core database if the details change due to parentage verification or parentage discovery using genetic testing (i.e. unknown to known). The field is not updated as a result of AE as all reporting of breed is in breed proportions.

The breed of a calf is noted when its birth is recorded. Its parentage determines its breed and this is derived from reproductive and mating information.

Breed 16ths shall be submitted with a 'Breed Type Code'. For the Breed Type Code details contact Support.NZAEL@DairyNZ.co.nz.

There is a Breed Type Code for Friesian but no Breed Type Code for Holstein. A 'Country of Origin Code' should also be provided. Friesian 16ths for animals originating from any country that is not NZ, Australia, or Great Britain is converted to Holstein 16ths as a part of the process for producing the AE pedigree extract.

8.2 Recording of parentage

HRPs shall specify an animal's parentage, both sire and genetic dam, if known, when animals are enrolled through the 'Animal Ancestry Event' interface. Only the CDP that is an HRP for 'Progeny Identification' can submit this event for an animal. HRPs shall include an 'Official Indicator Type Code' when they record an animal's parentage (sire and dam) to indicate the level of confidence in the parentage status for both sire and dam.

The sire, sire official indicator[3] and the dam are regulated and the data requirements for them are set out in the Herd Test Standard. Valid values are included in the DIGAD Interface Specifications.

The Official Indicator Type Code for the dam is non-regulated data and the valid values are included in the DIGAD Interface Specifications. The Official Indicator Type Code options for the dam, which are similar to those for the sire, include:

Null	Unknown
0	No mating recorded
3	DNA results in a negative parentage test
5	Not verified, uncertain, may be two or more sires, birth date does not meet the recommended range in the DairyNZ Sire Determination Methodology
7	DNA results indicated that parentage is probable, but no supporting meta data provided. DNA verification shall be performed using the ICAR requirements for single parent-offspring verifications as described in section 8.4.5.1 *.
9	Mating indicates that parent is possible as set out in the DairyNZ/NZAEL Sire Determination Methodology
10	For international animals where overseas evidence indicates that parent is possible

*Additional SNPs utilising the ICAR554 or ICBF800 SNP manifests can and are recommended to be used for parentage verification, and results must be prescribed using the ICAR parent discovery thresholds described in section 8.4.5.3 or the more restrictive thresholds described by the ICAR parentage guidelines in 8.4.5.1. Note the SNPs listed in section 8.4.5.4 should not be used. In the case where only a reduced subset of no less than 200 SNPs out of the ICAR554 SNPs are available, a parentage result using the methodology and SNP conflict thresholds described in 8.4.5.1 will also be accepted.

The SNP data, or the metadata (section 8.3) associated with the DNA testing used for an Official Indicator Type Code 7 entry shall be accessible to an audit if requested by a Certification Body or by NZAEL.

The database processing system can also derive the Official Indicator Type Code automatically when a calf record is processed. The processing system assigns an official indicator according to rules set out in the DairyNZ Sire Determination Methodology.

Once the record is established in the core database it is only updated:

- By an HRP if the details change via parentage verification or parentage discovery using genetic testing (i.e. unknown to known); or

[3] Referred to as the 'Sire Official Indicator Type' in the DIGAD Interface Specifications.

- By NZAEL reporting parentage following routine animal evaluation if parentage verification is not possible based on the information recorded. In such circumstances a code 7 shall be recorded as a code 3 or 5 until further DNA testing has been completed and the dam updated. Whether an HRP updates the parentage information in their propriety herd management system is at an HRP's discretion and not the purview of this standard.

8.3 Recording of SNP based DNA parentage testing

The value of DNA parentage testing is strongly influenced by the quality of the metadata provided with the DNA result. The type and basic details of the certified DNA result should be recorded, or available, to the DIGAD and include at a minimum for SNP exclusion based tests:

- AnimalDurableKey,
- TestDate,
- Certified data provider who submitted this result,
- TestType of parentage verification associated with the SNP panel or technology used (verification with the ISAG200, ICAR554, or ICBF800 panels, other accredited test, or other approaches investigated in future),
- Number of SNPs called in common in both sire and progeny
- Number of SNPs called in common in both dam and progeny
- Number of SNPs called in common in all of sire, dam, and progeny for mating test
- Sire result
- Dam result
- Mating test result
- Number of conflicting SNP - Sire
- Number of conflicting SNP - Dam
- Number of conflicting SNP - Mating test

A Certified Genotype Provider can submit a parentage result ("Parentage Test Event") to DIGAD for any animal regardless of who an HRP is for that animal. The parentage result may also be provided to the client submitting the sample from the animals for parentage testing via the Certified Data Provider – nominator for genetic testing.

8.4 SNP based parentage verification using exclusions (informative)

SNP based parentage verification can be computed with an exclusion-based process since every allele the offspring inherits must have come from either the sire or dam. The process tests every individual SNP locus that is called in common between:

1. an offspring and its putative sire (sire test),
2. an offspring and its putative dam (dam test),
3. an offspring and its putative sire and dam combined (trio or mating test).

Passing a SNP exclusion-based verification test does not guarantee that a nominated sire and offspring have a sire-offspring relationship, but only that the samples in question are similar enough that it is a possibility.

A given SNP is considered a pass for that locus when both the progeny and its putative parent (either sire or dam) are equal homozygous (i.e. AA). A fail at that locus occurs when a progeny and its putative parent are opposite homozygous (AA versus BB). In the case of opposite homozygosity, an offspring that is BB must have inherited a B allele from both parents, so a parent that is homozygous AA cannot be a candidate. If either progeny or parent are heterozygous at that location, the test at that locus is not informative.

The mating (or trio) test performs the same homozygous test as in the single parent case between the progeny and its putative sire and the progeny and its putative dam, however, it can also consider the cases where a sire and dam are equal homozygous, i.e. both AA. If the progeny is not also AA then this SNP locus is considered indicative of a parentage failure. Note that the mating test described by the ICAR parentage guidelines included verbatim in 8.4.5.2 states the progeny is a conflict only when the parents are homozygous and the progeny is heterozygous. The cases where the progeny is opposite homozygous (and the parents are homozygous) or the case where the parents are opposite homozygous and the progeny is heterozygous are not described by ICAR as a conflict.

Genotyping systems do exhibit errors, for example, a heterozygous (AB) animal may be incorrectly identified as AA or BB. Genotyping errors are infrequent, estimated as 0.1% of the calls, but that will represent 50 incorrectly called loci in a 50k genotyping array. Accordingly, a single locus mismatch is not sufficient to eliminate a progeny and putative parent from representing a parent-offspring relationship. Neither are a few mismatches. Incorrect parent offspring pairs are typically characterized by many mismatches.

For greatest accuracy, many SNPs should be used for parent verification. Genotype samples, especially historical ones, may originate from various chips (SNP content). These may have varying overlap with modern and future chips. An international standard exists for using the ISAG200 SNPs for parentage verification, however the small number of SNPs used on that panel means that there is a higher probability of there being insufficient mismatches to exclude a parent and putative offspring, even when they do not represent the true parent-offspring pair. This means that subsequent parent verification using a larger number of loci may not support parent-offspring pairs that were previously accepted.

From an industry perspective, genotyping as many animals as possible on the same chip content increases volumes for that platform and can reduce cost for the laboratories (which may or may not be passed on to the consumer). Utilising 50K (or higher) density chips for parentage means the same content can also be used for parentage verification and/or discovery. Low density “parentage only” panels are not useful for genomic evaluation purposes, and depending on the number of SNPs shared between the putative parent and offspring pair, may not produce the exclusion result that might have been obtained using SNPs from higher density panels.

Raw or actual genotype calls, not imputed calls, should normally be used in parent verification. Imputation is a process that alters missing genotype calls or fills in loci, and can be undertaken using pedigree only, population only, or both pedigree and population information. One exception to using imputed genotype calls is when these have been reconstructed, for example when DNA samples are no longer available for an ancestral sire or dam, but parent-verified genotyped offspring, or genotyped parents, are available for use in reconstruction.

8.4.1 Process for parentage discovery using exclusions (informative)

The process for discovering parentage is much the same as for verification, except that many offspring and putative parent pairs must be considered. Effectively, discovery involves attempting parentage verification between a progeny and all potential sires (and/or all potential dams and trios) to find matches. This search can be performed against every known genotyped sire, or ideally against a smaller list of potential sires (such as those used in natural service matings after AI), or the list of active bulls in an AI program. The same principles apply to dam discovery, the test can be undertaken against all the dams in the herd, or in the same embryo transfer program.

Parentage discovery can be performed separately to search for potentially matching sires and dams, and then all possible combinations of the potentially matching parents can be submitted to the verification process in order to a) double check the result and b) perform the additional trio (or mating) test between progeny-sire-dam.

Knowing the birth year and pedigree recorded sex of the animals in question improves the accuracy and efficiency of parentage discovery as then it is possible to automatically exclude impossible matches and prevent progeny or siblings being assigned as potential parents. With the increase of embryo transfer in particular, incidences of identical twins, while rare, can occur dozens of times for population sizes of six figures or more.

8.4.2 SNP exclusion thresholds (informative)

There are two types of exclusion thresholds used for both the single parent and trio mating tests. In both tests, there is a separate absolute exclusion threshold for the number of allowable conflicting SNPs, and a proportion threshold, measured as the number of conflicting loci divided by the total number of informative loci. The number of informative loci is the number of loci that were called in common between the samples involved in that particular test.

For general purpose (non-ICAR) verification and discovery, an individual proportion threshold of 1% and a trio conflict threshold of 1.5% are used.

8.4.3 Data requirements

Raw SNP chip genotype data, i.e. not imputed, or otherwise reconstructed genotype calls, shall be used for parentage activities. For general genotype data requirements see the relevant section under Genetic Testing. SNP data may include loci that are not mapped or are present on X or Y (sex) chromosomes. Only loci inherited

in an autosomal manner and which pass quality control procedures should be used in parent discovery or parent verification Only the X and Y chromosome loci in the pseudo-autosomal regions are inherited in an autosomal manner but for simplicity the X and Y chromosome SNPs should be excluded from parentage analyses. The following SNP panels are supported for parent verifications for code 7 Official Indicator Type Code assignment:

- ISAG200 – see https://interbull.org/ib/pse_parentage_verification_snps or https://wiki.interbull.org/public/PV_SNPs?action=AttachFile&do=get&target=GenoEx-PSE_PV_SNP_List_2017.csv
- ICAR554 – see https://www.icar.org/wp-content/uploads/2020/09/ICAR-554-SNP-List-for-Parentage-Discovery.xlsx
- ICBF800 – see https://pmc.ncbi.nlm.nih.gov/articles/PMC5862794/

8.4.4 NZAEL reported parentage (informative)

NZAEL is the custodian of the core database and the manager of DIGAD which stores the SNP genotype data used for AE. NZAEL reported parentage means an animal's parentage testing results have been generated as a part of routine quality control processes for AE using the raw (not imputed) SNP genotype data. This test is performed using all available quality-controlled SNPs and can be reported back to HRPs through the same process by which AE results are reported. The primary objective of NZAEL based parentage is to provide accurate genetic evaluation by eliminating, or correcting, pedigree errors or excluding problematic genotypes. Reporting these results back to HRPs enables transparency about which genotypes, or pedigree records, may be excluded from the evaluation due to failing parentage. NZAEL reported parentage, performed as part of the routine AE is more accurate than the ICAR-based parentage described in 8.4.5 due to its use of all available commonly-called SNPs, in comparison to the more limited use of ISAG200 SNPs for ICAR certified parentage. This means some animals may pass ICAR tests but fail the more rigorous NZAEL parentage tests, as also described by the Council on Dairy Cattle Breeding (CDCB) (see the policy https://redmine.uscdcb.com/attachments/download/8825/CDCB%20policy%20-%20ISAG-based%20SNP%20parentage%20tests%20and%20certificates.pdf). According to the Irish Cattle Breeding Federation (ICBF), a minimum of 500 SNPs is needed to verify animals with <= 1% miscordance rate (see article - https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2018.00084/full).

8.4.5 ICAR accredited parentage (informative)

ICAR provides guidelines for performing SNP based parentage verification and discovery, outlined - https://www.icar.org/wp-content/uploads/2023/07/ICAR-Guidelines-for-Parentage-Verification-and-Parentage-Discovery-September-2022.pdf. These include which SNPs should be used, the number of SNPs required for an acceptable result, and the number of conflicting SNPs which constitute a failure.

The ICAR guidelines have and will continue to update over time as newer technologies and algorithms develop.

For reference the currently published ICAR guidelines have been copied verbatim into subsections 8.4.5.1-8.4.5.4.

8.4.5.1 ICAR guidelines for verification of a single parent-offspring pair:

Step 1: Conduct a separate verification for each combination of the animal with its recorded sire and/or dam with a SNP genotype. The informative SNP are those for which the animal and reported parent are both homozygous and a conflict is considered when they are each homozygous for a different allele for any informative SNP. Based on the minimum criteria of 185 SNP available for the animal and each parent, the minimum number of common SNP available for verifying each animal-parent combination is 175 (i.e.: 195 – (2 x (195-185)) = 175).

For this step, the following rules apply for assigning the parentage verification status:

- Number of mismatches/SNP conflicts: 0 – 2 => Parent Accepted
- Number of mismatches/SNP conflicts: 3 – 5 => Parent Doubtful
- Number of mismatches/SNP conflicts: >5 => Parent Excluded

8.4.5.2 ICAR guidelines for mating verification of a sire-dam-offspring trio:

Step 2: In the case that both sire and dam have a status of “Parent Accepted” from Step 1, verify that the combination of those parents is acceptable. In this case the informative SNP are those for which both verified parents are homozygous and the progeny is heterozygous. A conflict exists when the parents are homozygous for the same allele at any informative SNP while the progeny is heterozygous. In this case, the minimum number of common SNP available is 165 (195 – (3 x (195-185)) = 165).

For this step, the following rules apply for confirming the parentage verification status for the combination of verified parents:

- Number of mismatches/SNP conflicts: 0 – 3 => Mating Accepted
- Number of mismatches/SNP conflicts: 4 – 7 => Mating Doubtful
- Number of mismatches/SNP conflicts: >7 => Mating Excluded

8.4.5.3 ICAR guidelines for parentage discovery using SNPs

Organisations carrying out parentage discovery services must implement quality assurance procedures that ensure the following:

- That a discovered parent is older than the animal and, in fact, not an offspring.
- That a discovered parent is of the appropriate sex such that sires are male and dams are female.
- That procedures must include methods for identifying and listing genetically identical animals. While this can be done as a separate processing step, the ICAR accreditation for parentage discovery will test that such animals are identified within the processing procedures such that a discovered parent is also reported as having known genetically identical siblings.

Given that genotyping SNP chips actively being used in cattle populations globally have a varying number of the 554 SNP defined for inclusion in the GenoEx-PSE service, parentage discovery results must be based on a percentage of SNP available between the animal and any potential parent being considered.

The following is recommended for assigning the parent discovery status:

Step 1: In separate processes, attempt to discover either the sire (i.e.: the male older than the animal with the fewest conflicts) or dam (i.e.: the female older than the animal with the fewest conflicts) of the animal based on SNP genotypes available. Based on the minimum criteria for each SNP genotype to be included, as outlined in point 3 above, the minimum number of common SNP between the animal and each candidate parent must be 400.

For this step, the following rules apply for assigning the status of each parent discovered:

- Percentage of 400+ common SNP with a conflict: 0 to ≤0.5% => Parent Discovered
- Percentage of 400+ common SNP with a conflict: >0.5 to ≤2.0% => Parent Possible
- Percentage of 400+ common SNP with a conflict: >2.0% => Parent Excluded

For animals for which the discovery process led to all qualifying results as “Parent Excluded”, the discovery conclusion for that animal is considered as Parent Not Found. For animals that do not have enough SNP available and/or are found during the discovery process that there is no candidate parent with at least 400 SNP in common, the discovery conclusion outcome should be Parent Not Found instead of any of Discovered, Possible or Excluded as outlined above.

Step 2: In the case that an animal has both a sire and dam with a successful status of “Parent Discovered” from Step 1 or resulting from parentage verification, this parent combination must also be checked.

For this step, the following rules apply for assigning the status of the combination of parents discovered, which requires the trio of the animal with both of the candidate parents after discovery to have at least 400 SNP in common:

- Percentage of 400+ common SNP with a conflict: 0 to ≤1.0% => Mating Confirmed
- Percentage of 400+ common SNP with a conflict: >1.0 to ≤4.0% => Mating Possible
- Percentage of 400+ common SNP with a conflict: >4.0% => Mating Excluded

For trio combination of animal with its candidate parents that do not have at least 400 SNP in common, the discovery conclusion outcome for the mating should be Mating Not Checked instead of any of Confirmed, Possible or Excluded as outlined above.

8.4.5.4 ISAG200 SNPs to be excluded from use in parent verification or discovery

SNP Name	ISAG Group	Reason for Exclusion
ARS-USMARC-Parent-DQ837645-rs29015870	Core	Clustering issues*
ARS-USMARC-Parent-DQ786766-rs29012070	Core	Clustering issues*
ARS-BFGL-NGS-76191	Backup	Clustering issues*
BTA-100621-no-rs	Backup	Clustering issues*
ARS-BFGL-NGS-99210	Backup	Tri-allelic**

*McClure et al. (2015)

** - Based on sequence validation to be specifically problematic with bead chips

8.4.5.5 Supporting documents (informative)

The following documents provide further detailed information on the ICAR accreditation process for DNA Data Interpretation Centres (note over time these may become out of date and links may become broken – presently the latest information can be found at this link - https://www.icar.org/index.php/certifications/dna-certifications/certification-and-accreditation-of-dna-genetic-laboratories/two-new-dna-based-services/dna-data-interpretation-centres/):

- Standard Operating Procedure for the submission for the ICAR accreditation of as DNA Data Interpretation Centres
- Application form for ICAR Accreditation of DNA Centres
- Applicant's Guide for the Parentage Analysis Accreditation for DNA Data Interpretation Centres
- ICAR Guidelines for Parentage Verification and Parentage Discovery based on SNP
- Distribution of SNP by Chromosome Approved by ICAR for Parentage Verification and Exchange in GenoEx-PSE
- List of SNP Approved by ICAR for Parentage Verification and Exchange in GenoEx-PSE

ICAR accreditation is also a requirement to participate in the ICAR international Genotype Exchange service (GenoEx). Additional information: explanatory background to GenoEx and GenoEx website.

8.4.6 Automatic incorporation of future ICAR guidelines for parentage verification and discovery

Any standards, guidelines, or recommendations officially adopted and promulgated by the International Committee for Animal Recording (ICAR) with respect to SNP based parentage verification or discovery shall be automatically incorporated as accredited parentage methods for the purposes of reporting a parentage result as described in 8.2. This incorporation shall take effect immediately upon formal acceptance and publication by ICAR, without necessitating further review or amendment to this standard.

8.4.7 Familial or reconstructed DNA, and probabilistic based parentage (informative)

Familial and/or reconstructed DNA profiles may sometimes be used to assist in identifying and correcting pedigree errors on historical animals who do not otherwise have a DNA verified parentage result recorded in DIGAD (i.e. a code 7). Reconstructed DNA profiles may be generated as part of AE processes; however, these will normally be used for improving genetic evaluation accuracy rather than certifying parentage status.

Probabilistic based parentage approaches may be evaluated in future when a high proportion of the national herd is genotyped, to complement or replace the exclusion-based parentage approach.

Familial DNA, for instance, based on haplotype identification and matching, or identification of maternal grand sires (MGS) may be undertaken, once a sufficient volume of genotype data allows it, to periodically examine the pedigree for errors and obtain retrospective corrections.

8.4.8 Updating incorrect parentage in DIGAD using genotype data

As a consequence of SNP based parentage verification (or discovery), pedigree records within DIGAD may be updated after confirmation of a valid match.

Parentage records that fail the verification process shall be reported back to HRPs so they can request alternatives from farmers or resolve genotype sample issues. See section 12 for more information on genotype sample issues and identification).

By default, genotypes of individuals which fail parentage testing during routine NZAEL processing for AE will not be included in any genomic evaluation.

9 Breed Proportions (informative)

Determining an animal's pedigree-based breed proportions is a natural consequence of having accurate parentage information, provided the breed composition of their ancestral founders is known and accurate. For the purposes of a recorded pedigree, a "founder" is any animal where one or both parents are unknown. The current pedigree includes many such animals, some with a known sire and unknown dam, some with a known dam and unknown sire, and some with an unknown sire and unknown dam. There is considerable value in identifying the correct parentage of these animals, particularly if they have extensive phenotypic records, as it automatically informs the breed proportion as well.

Farmers currently nominate breed proportions in the breed 16ths field when they enrol animals. In AE, NZAEL estimates the breed proportion of every animal in decimal form, as half of its known sire's breed proportion, plus half of its known dam's breed proportion. Breed proportions are computed in this way from the oldest founder animals in the pedigree to the youngest. The farmer-entered breed 16ths for animals with known parentage are, therefore, not used directly. However, they may be helpful in identifying and cross-referencing genotype sample information when there are sample issues or other quality control problems, such as failed parentage.

For animals where one parent is known and one unknown, the breed 16ths of the animal and the computed breed proportion of the known parent are used to estimate the breed proportion of the unknown parent, before continuing down the pedigree to compute the breed proportions for any descendants.

For animals where both parents are unknown, the breed 16ths of the animal, as provided by the farmer, is converted to decimal form and used directly.

Consistency checks are performed to ensure that an animal's breed proportions all sum to one, and that it does not have a specific breed proportion that is less than half of either of its known parents' breed proportions for that same breed.

An additional complication is that DIGAD did not historically distinguish Holstein from Friesian animals, and therefore did not store Holstein breed proportions. In recent times, the Holstein breed proportion has been routinely computed from founders and used in AE. An animal's country of origin is recorded into the DIGAD. For founder animals recorded as Friesian, but with an overseas origin, the recorded Friesian breed proportion is converted to Holstein. Holstein breed proportions are then inherited by their descendants from the founder down calculation.

The breed 16ths and country of origin recorded in DIGAD are utilised in various forms to help guide the assignment of breed proportion in founder animals for use in AE. The exact methods by which founders are assigned breed proportions, and in particular the manner in which Holstein breed proportions are assigned to founders, may change over time if new information or methods become available to further improve the overall accuracy of genetic evaluation.

In a multibreed evaluation accurate estimation of breed effects is critical in accurately ranking selection candidates across breeds. Phenotyped animals (or parents of phenotyped progeny) with incorrect breed proportions may cause inaccurate predictions of breed effects. This will bias the individual animal that has the incorrect breed assignment, the nationally estimated breed effect, and all animals in the evaluation.

10 Genetic Variants and Major Genes (+ve and -ve) (informative)

Numerous genetic defects or SNPs may be of interest due to their potential impact on animal welfare (e.g., horned/polled, cancer, heat stroke, or congestive heart failure risk), production or value-add (e.g., A2), or other serious recessive disorders affecting an animal's life expectancy or productivity.

SNPs associated with patented or licensed genetic defects generally are not reported with raw genotype data for genetic evaluations. Instead, the laboratories responsible for validating genetic tests generate validated defect reports and return them directly to the relevant HRPs or end users.

The list of genetic defects continues to grow as new defects are identified through genomic analyses. The importance of new genetic defects increases when a bull carrying an undiscovered defect is widely used. In contrast, avoiding carrier animals in bull selection can reduce the importance of testing for previously significant defects over time.

DIGAD should store the status of tested animals for industry-important defects and major genes, including carriers not afflicted by recessive conditions. Only validated genetic test results shall reported by the laboratories.

10.1 Recording genetic tests

Genetic tests are recorded as part of the 'Animal Event' interface using the 'Locus Type Code' the 'Allele Type Code'.

Ensuring appropriate animal data for AE is the highest priority. To prevent evaluation bias, carriers of certain genetic variants require exclusion of their (extreme) phenotypes from AE.

In situations where industry stakeholders confidentially hold specifics of a tested variant for commercial reasons, sharing a list of afflicted animals with phenotypes requiring exclusion from AE (to prevent bias) could serve as an alternative.

10.2 List of genetic variants or major genes

The defect status of all enroled bulls should be reported as: afflicted, carrier, or free, for any required genetic tests (see Table 1). It is the responsibility of the laboratory to only report results from validated tests, i.e. tests that have been validated against legitimate afflicted, carrier, and/or free animals as appropriate.

For all other tests and animals, any available results should be reported when they are available.

Table 1. Genetic variants reported by laboratories.

Recorded genetic test	Group	Reporting status
Beta-casein A2	Milk	Recommended
Beta-lactoglobulin	Milk	Recommended
Kappa-casein	Milk	Recommended
Horned/polled	Unwanted	Recommended
BLAD – bovine leukocyte adhesion deficiency	Unwanted	Recommended
Mule foot	Unwanted	Recommended
DUMPS – deficiency of uridine monophosphate synthase	Unwanted	Recommended
CVM – complex vertebral malformation	Unwanted	Recommended
Factor XI - blood clotting	Unwanted	Recommended

Recorded genetic test	Group	Reporting status
CIT – Citrullinemia	Unwanted	Recommended
Brachyspina syndrome	Unwanted	Recommended
Cholesterol deficiency	Unwanted	Recommended
Small calf syndrome (liveweight evaluation)	Unwanted	Recommended
Embryonic lethals (fertility evaluation)	Lethal	Recommended
Gestation length (Amoret)	Unwanted	Recommended
Hairy calf	Unwanted	Recommended

10.3 Terms of reference for tests with required reporting status for bull enrolment (informative)

The following factors influence whether a variant should be recorded when enrolling a bull:

- the variant is internationally reported and/or represented in sire catalogues;
- if there is a direct phenotypic effect on the afflicted animal;
- it will bias AE if its phenotype data (or that of its descendants) is not excluded, (e.g., small calf syndrome, embryonic lethals in fertility, gestation length);
- if defects (or variants) are dominant effects; and
- availability for a commercial genetic test for that variant at genotyping laboratories.

Periodically, the list of genetic variants should be reviewed by a committee of industry stakeholders to ascertain which, if any, variants should be nominated to have required reporting status.

11 Genetic Testing (informative)

The primary objective of genetic testing, or genotyping, is to improve animal evaluation (AE) by providing accurate and comprehensive genetic data to improve AE predictions.

Farmers have the option to choose from certified laboratories and approved products and whether they meet farmers' requirements for:

- parentage, and
- genotype profiles that are suitable for genomic evaluations and inclusion by New Zealand Animal Evaluation Limited (NZAEL) for AE purposes.

This allows farmers to have confidence when selecting genetic testing providers and products from listed providers and they will also know which data will be sent to the DIGAD.

As AE and genetic testing technology evolves over time, the capability to use data in new ways may emerge. AE will continually adapt to incorporate the best data in optimal ways, ensuring that the evaluation process remains robust and up-to-date with advancements in technology and research.

11.1 See Appendix 3 for more information on the certified data providers relating to this standard. Certified Data Provider – nominator for genetic testing (informative)

There are many layers to the genotyping process. The process begins on farm with the collection of animal tissue samples, typically using a tissue sample unit (TSU). Ideally samples should be collected while simultaneously scanning the TSU barcode and the electronic identification (EID) tag of the animal. This electronic association of TSU barcodes with EIDs minimises the need for handwritten recording of sample metadata and reduces the risk of errors.

Depending on the laboratory selected by the client, the tissue samples and associated metadata on the laboratory sample submission sheet can be sent:

- directly to the laboratory, where the laboratory is both a CDP – nominator for genetic testing and providing genotype laboratory services; or
- to a CDP – nominator for genetic testing[4] who will liaise with NZAEL, and the client as necessary. They will ensure the samples for genetic testing have the correct formats for submitting data to the nominated genotyping laboratory and each sample is referenced to the animal durable key before the samples are processed by the laboratory. Laboratories test the samples to generate DNA information and submit these data and associated metadata directly to NZAEL, or their nominee, and to the Certified Data Provider – nominator for genetic testing if the client requests it.

11.1.1 Standard for Certified Data Provider – nominator for genetic testing (informative)

The role of the CDP – nominator for genetic testing can be undertaken by a laboratory as part of the genotyping services they provide, or by a person or organisation independent of the genotyping laboratory.

The Standard for the Certified Data Provider – nominator for genetic testing can be found here.

11.1.2 Test request sheet for sample submission

Certified Genotype Data Providers shall deliver genotype data to NZAEL, or their nominee, with the necessary files to cross-reference between delivered sample IDs and the associated DIGAD animal IDs, e.g., the Animal Durable Code. Individual laboratories may be able to perform this cross-referencing themselves as part of their delivery of genotyping services, either through their own system and/or through connecting directly with DIGAD (i.e. they undertake the Certified Data Provider – nominator for genetic testing role).

Certified Data Provider – nominator for genetic testing shall submit a sample submission sheet or test request sheet with the samples for analysis to enable the cross-referencing of sample IDs. The sample submission sheet may differ for each approved laboratory depending on their individual system requirements (see list of example data in Table 2), but in all cases the submission sheet will require a minimum of five fields: the Sample ID barcode or descriptor, the Animal Durable Code associated with the animal from which the sample was taken, an animal identifier, the Certified Genotype Data Provider, and the laboratory product on which the samples will be tested. Where a laboratory sample submission sheet is a required field, and the data are stored in the DIGAD, the data for the required fields shall be sourced from DIGAD unless the genotyping laboratory has access to the data via an HRP. The CDP – nominator for genetic testing shall resolve any conflicts between data sourced from DIGAD and data sourced from an HRP.

The sample submission sheet or test request sheet fields shown in Table 2 are an example of the data required by NZAEL and these may also be required by the laboratory. Sample submission sheets and instructions are available via the approved genotyping laboratory link.

[4] Note: the CDCB nomenclature is “genomic nominators”

Table 2. Example of the data required in a test request/sample submission sheet.

Column Header Name	Description	Status
Sample ID	TSU barcode or hair card number (or equivalent)	Required
AnimalDurableCode	DIGAD AnimalDurableCode – i.e. the AnimalDurableKey, with suffix 'DK' left padded with 0's to make 12 digits, e.g., DK000012345678.	Required
Birth ID	On-farm identifier for this animal, e.g. Birth ID	Required – if EID Not included
EID	Electronic NAIT ear tag ID	Required – if Birth ID not included
Sex	One letter sex code, e.g. • M – male • F – female	Recommended
Breed	e.g., Uniform Breed Code or 3 letter breed code	Recommended
Date of birth	dd/mm/yyyy	Recommended
Sample type	One letter type code, e.g. • H – hair • S – semen • T – tissue (muscle sample) • G – tissue sampling unit (TSU), e.g. Allflex • D – done (laboratory already has sample) • U – profile build	Recommended
Laboratory (i.e. certified genotype provider)	Accredited laboratory code. See the DIGAD Interface Specifications for the code or the register for Certified Genotype Providers.	Required
Laboratory product	Product used to analyse the samples – specific to each lab. See the register for Certified Genotype Providers.	Required
Participant code	Participant code, i.e. the owner of the animal	Recommended

11.2 Certified genotype data provider

Certified genotype data providers shall have ISO 17025 laboratory accreditation and be able to supply standardised genotype data files and the associated metadata to NZAEL via FTP (or equivalent webserver (e.g. AWS S3)). Genotype data providers are not required to report parentage results to the DIGAD. Providing genotype data to NZAEL enables the data to be used to improve AE; however, it does not guarantee genomic

predictions will be generated for the animals whose genotype data is provided if the data is determined by NZAEL to be unsuitable for AE purposes, for example via its genotype quality controls.

Certified genotype data providers shall deliver genotype data to NZAEL (or their nominee) with samples labelled using a sample identifier which is relevant to their internal Laboratory Information Management System (LIMS). This enables the genotype data provider to link any queries relating to individual sample results back to their internal processes to identify any potential issues.

Genotype data should be provided with an associated metadata file (of which several forms are accepted) that links the provided sample IDs back to the Animal Durable Code of the genotyped individuals in question.

See the register of certified genotype data providers and their products.

11.2.1 Genotype laboratory accreditation

- Genotype laboratories shall be at least accredited for ISO17025, e.g. https://www.standards.govt.nz/shop/nzs-isoiec-170252018 and New Zealand based laboratories shall hold this accreditation via IANZ.
- Genotype data shall be reported with SNP locations, either according to accepted standard reference genomes, Illumina SNP names, RS numbers, or an equivalent accepted naming convention (as in https://www.icar.org/wp-content/uploads/2015/08/Annex-IV-guidelines-SNPs.pdf).
- Genotype data providers shall provide an FTP server (or equivalent file share mechanism) where they will place genotype and metadata files. Each genotype data submission shall be placed in its own uniquely named folder.

11.2.2 Genotype and metadata file formats

File delivery consists of a genotype data file ('FinalReport' file) and a corresponding sample metadata file which can exist in one of several different formats (see Table 3).

Genotype data should be delivered with the Sample ID relevant to the genotype laboratory in the final report, e.g. this may be the original sample barcode or may be an internal laboratory sample key associated with the sample in the laboratory's LIMS.

The associated sample metadata file shall have at least the two fields corresponding to the Sample ID and its cross-reference to the Animal ID field (AnimalDurableCode).

Table 3. Example file formats for submission sheet data and metadata.

File	Description	Status
Submission sheet file formats (example only)		
TestRequest.csv	An example test submission sheet provided for reference. It is expected individual genotype laboratory providers will have their own submission sheet requirements and will manage that process with their customers directly.	Optional – not required alongside genotype data delivery
Genotype data and manifest file formats		
*_FinalReport.zip	Genotype data reported in standard GenomeStudio Final Report file format. The filename preceding "_FinalReport.zip" can contain any relevant information, e.g. YYYYMMDD_CHIP_FinalReport.zip. It shall unzip to a .txt file of the same name, e.g. YYYYMMDD_CHIP_FinalReport.txt.	Required – shall be provided with every genotype data delivery
SNP_Map.zip (or equivalent)	Chip manifest for the SNPs included in the final report file including chromosome, position, and any other relevant information.	Recommended – shall be provided at least once for each product. Can be supplied with

File	Description	Status
		every data delivery or supplied only when requested i.e. when the chip manifest changes
Sample-Animal linking metadata file formats (in order of preference). At least one of the possible metadata files shall be provided with every genotype data delivery, e.g. either a AnimalDetails.zip file or SampleSheet.zip file.		
*_AnimalDetails.zip	AnimalDetails file, e.g. YYYYMMDD_AnimalDetails.zip	(One of) required
*_SampleSheet.zip	SampleSheet file , e.g. YYYYMMDD_SampleSheet.zip	(One of) required
Other useful sample metadata files of note		
Sample_Map.zip	Mapping of samples to physical chip positions, generally applicable to Illumina platforms	

The genotype data must be delivered in the standard Illumina Genome Studio final report text format with as a minimum data columns including SNP name, Sample ID, AB – Allele1, AB – Allele2, and GC score or an agreed upon functionally equivalent metric on the confidence of each SNP call. Ideally TOP and FORWARD alleles should be included as well. The file should be zipped (it can be saved in .zip, txt.gz, or txt.bz2 file formats, but shall unzip to a file with filename ending in “FinalReport.txt”). The prefix of the filename is flexible but might include date, order number, batch number, or other metadata. Genotypes from Thermo-Fisher (Affymetrix) genotyping systems are also acceptable, provided the calls meet quality control standards, and the files meet Illumina Final Report file naming and content conventions.

A sample sheet shall be provided with the final report file to link the provided sample IDs with the associated Animal Durable Code. Standard formats, any one of which are acceptable, are described in section 11.2.2.1.

Table 4. Example final report file showing the required data. This contains genotype data submitted by the laboratory.

Column Header	Description	Status
SNP Name	Illumina SNP naming conventions	Required
Sample ID	TSU barcode or hair card number or equivalent, e.g., internal laboratory LIMS key	Required
Allele1 – AB		Required
Allele2 – AB		Required
Allele1 – Forward		Required
Allele2 – Forward		Required

Column Header	Description	Status
Allele1 – Top		Required
Allele2 – Top		Required
GC Score	On Affymetrix platforms this field may be called a "Confidence"	Required – where available, e.g. Illumina platform
LogR Ratio		Optional

11.2.2.1 Genotype sample metadata file examples

There are several possible sample metadata file formats (examples below) (note, the order of the included fields (columns) does not matter as the column header labels are used to process the files).

AnimalDetails.zip

Table 5 shows the data included in an AnimalDetails.zip with a comma separated (csv) file format – this format is routinely generated in beef cattle genotyping pipelines. Additional columns not listed in Table 5 can be included without issue.

Table 5. Example genotype metadata file format – AnimalDetails.zip in comma separated (csv) file format.

Column Header	Description	Status
Barcode	SHALL align exactly with the Sample IDs provided in the Final_Report file, typically the TSU barcode or hair card number or internal laboratory sample ID.	Required
Animal ID	DIGAD AnimalDurableCode – i.e. the AnimalDurableKey, with suffix 'DK' left padded with 0's to make 12 digits, e.g., DK000012345678.	Required
Barcode 2	An optional secondary ID that can be associated with the sample	Optional

SampleSheet.zip

If a SampleSheet.zip with a comma separated value (.csv) file format is used a header row with field names should exist following a [Header] section and immediately after the [Data] line in the file. The file should be comma delimited e.g. the start of the file should look something like the following:

[Header],,,,,,,,,,,,,,,,,,,,,,

Investigator Name,NAME,,,,,,,,,,,,,,,,,,,,

Project Name,PROJECT,,,,,,,,,,,,,,,,,,,,

Experiment Name,EXPERIMENT_NAME,,,,,,,,,,,,,,,,,,,,

Date,4/24/2023 10:16,,,,,,,,,,,,,,,,,,,,

[Manifests],,,,,,,,,,,,,,,,,,,,

Axiom_VM2.r1,,,,,,,,,,,,,,,,,,,,

[Data],,,,,,,,,,,,,,,,,,,,

Sample_ID,Sample_Plate,Sample_Name,Project,AMP_Plate,Well in AMP Plate,SentrixBarcode_A,SentrixPosition_A,Scanner,Date_Scan,Replicate,Parent1,Parent2,Gender,Sample_Source,Study,Subclient,Well on DNA plate,Sex,Family Info,Other_Name,Sire,Dam,Comment

23083034,20230316_B24,DK000020851491,GenVis,20230316_B24,P12,GV4477344204,R16C12,,,,,female,Tissue,,,,,,,,

Any other columns not listed can exist and will be ignored.

Table 6. Example genotype metadata file format – SampleSheet.zip with a comma separated value (.csv) file format.

Column Header	Description	Status
Sample_ID	**SHALL align exactly with the Sample IDs provided in the Final Report file**, typically the TSU barcode or hair card number or internal laboratory sample ID.	Required
Sample_Name	DIGAD AnimalDurableKey – i.e. “DK” left padded with 0’s to make 12 digits, e.g. DK000012345678.	Required

Sample map file

A sample map file (Sample_Map.zip), which is tab delimited with columns, may be included alongside each genotype data submission (see Table 7).

Table 7. Fields included in a sample map file.

Column Header	Description	Status
Index	Unused	Required
Name	TSU barcode or hair card number or equivalent – should align with “Sample ID” in Test Request file	Required
ID	Unique Animal ID that returned sample data will be labelled with, e.g., AnimalDurableCode. Shall align with the IDs of the samples included in the Final Report file.	Required
Gender	One of “Male”, “Female”, or “Unknown”	Required
Plate	e.g., “346244”	Required
Well	e.g., “F7”	Required
Group	Can be left blank	Required
Parent1	Can be left blank, otherwise should be Animal ID (AnimalDurableKey)	Required
Parent2	Can be left blank, otherwise should be Animal ID (AnimalDurableKey)	Required
Replicate	Can be left blank	Required

Column Header	Description	Status
SentrixPosition	e.g. “205685290017_R17C03”	Required

11.2.3 Quality Assurance and Quality Control (QA/QC) reporting of submitted genotype data (informative)

Scorecard or QA/QC reports can be generated for each genotype data submission and returned to the certified genotype data provider and/or genotyping laboratory.

11.3 SNP array validation for animal evaluation

A variety of SNP chip arrays exist in the market. Even amongst “50K” products, which are SNP chips capable of sampling 50,000 SNPs per sample, there is often much less than 100% overlap between genotyping service providers. There are also multiple SNP array technologies and different SNP naming conventions. Therefore, new SNP arrays that providers want to use for routine genotyping shall be validated for quality, concordance, and overlap compared to the existing genotype data stored in the genomic database.

Although any segregating SNP can add value for parentage verification or discovery, not all SNPs add value to AE. The SNPs included in the evaluation may differ by trait and may change over time as part of ongoing efforts to increase the accuracy of genomic evaluation predictions and as part of the development of new SNP chip arrays.

To validate a new SNP array at least ninety genotyped samples should be submitted from animals representing the dominant NZ dairy breeds. At least ten samples shall have a previously genotyped sample already stored in, or accessible to DIGAD. If the SNP array submitted for validation has not yet been validated by the CDCB[5] (see list https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/REFERENCES#Ref162), then at least thirty samples must be submitted from animals with a previously genotyped sample in or accessible to DIGAD. In either case the remaining samples of the ninety shall be progeny of animals who have a genotype already stored in or accessible to DIGAD. Note that these are minimum requirements and if any issues are identified on the submitted samples, then additional samples may be requested in order to validate the array for use in AE.

11.4 Low-pass sequence data

Some genotype data providers are now offering low-pass sequencing as an alternative to SNP chip genotyping. This cost-effective method enables whole-genome sequencing at low coverage. However, the accuracy of low-pass sequencing results depends significantly on the depth of sequencing and the haplotype data used for assembly, which can vary extensively by breed and country (i.e. the reference population).

Unlike SNP chips, which rely on a predefined set of variants, low-pass sequencing offers a broader view of the genome. Nonetheless, SNP chip data often exhibit greater reproducibility and accuracy for the variants they are designed to detect.

Current genotype data formats, equivalent to an approved array manifest, do not sufficiently guarantee the quality of low-pass sequence data for inclusion in AE. Given the variation in accuracy by breed and reference population, developing appropriate metrics for using this data in productionised AE is the subject of ongoing research.

All delivered samples, even those meeting the acceptable data format criteria, will undergo rigorous genotype quality control checks as detailed in Section 12. These checks shall occur before any potential use in AE, including routine parentage, using a broad set of SNP. Low-pass sequence data products delivered in the currently accepted genotype data formats on a manifest equivalent to an approved array manifest are not sufficient on their own for low-pass sequence data to be acceptable quality for inclusion in AE.

[5] CDCB validated SNP arrays:
https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/REFERENCES#Ref162

As the technology and associated production implementation methodologies mature, the use of low-pass sequence data in routine AE will be revisited, ensuring that the most accurate and comprehensive data informs AEs.

NZAEL may use low-pass sequence data as part of research and development of AE.

11.5 Non-routine historical and overseas data

Provision of genotype data of either historical or non-routine (e.g. overseas) origin can proceed provided the data is delivered in one of the accepted genotype data formats or in a format suitable for AE. In situations where overseas genotype data is related to AE enrolment of a bull, if the internationally provided genotype is not sufficiently similar to the currently approved genotype data products utilised in AE, re-genotyping of the animal at an approved genotyping laboratory using a genotype product suitable for AE is required. Non-routine genotype data deliveries may be evaluated and accepted by NZAEL on a case-by-case basis (potentially at a cost), however, this does not guarantee the use of that data in AE, and all such data will always be subject to the same stringent genotype quality controls described in section 12. Likewise, provision of non-routine pedigree data can proceed for utilisation in AE provided it can be delivered in a data format which is suitable for AE processing.

12 Genotype Quality Control

Genotype samples undergo various quality control steps prior to being used in parentage analysis and AE (see Table 8). Failing a test means the genotype will be excluded from use in AE.

Table 8. Quality control tests used for genotype samples.

Quality control test	Passing threshold	Remedy if failing
Call rate	>=0.95	Resample required
Similarity (concordance of duplicate samples)	>=0.99	Retraction of bad duplicate(s). Correct sample can sometimes be determined via parentage and sex prediction
Sex prediction (X+Y chromosome)	Predictions of sex based on X+ Y loci match pedigree sex. Mismatch in both is a failure. Mismatch in one or the other requires manual verification.	Complete failures (mismatch in both X and Y prediction against pedigree sex) requires the genotype data provider to verify with the farmer that the pedigree animal sex is recorded correctly. If it is, generally a resample would then be required. Manual verification requires manually overriding the predicted sample sex.
Parentage (NZAEL AE result only)	• <= 1% of informative SNPs conflict for individual sire/dam tests • <=1.5% of informative SNPs conflicts for trio (mating) tests	Failure indicates one of the involved samples does not belong to the correct animal, or that the recorded pedigree is incorrect. Remedies include: • Verifying if the sire and/or dam are passing to other progeny (which indicates their samples are OK) • Performing genotype based parentage discovery to find a verifying alternative match • Correcting the recorded pedigree • If the sample does not match to any of the likely parents (who are genotyped) then a sample mismatch may have occurred, and

Quality control test	Passing threshold	Remedy if failing
		resampling the animal is the next best step.

12.1 Call rate

The call rate of any given genotype sample is the ratio of SNPs that were successfully called (i.e. measured) for a given sample, divided by the number of SNPs that were available on the chip that the genotype was sampled on. Call rate is typically the first quality control criterion checked when processing genotype samples. Samples that do not meet the required call rate are invalidated from further analysis.

Deterioration of tissue samples in transit, poor quality or contaminated samples, or poor laboratory handling practices may lead to lower call rates. Genotype concordance rates between repeated genotype analyses fall below 99% when the call rate drops below 0.9 (Cooper et al., 2013[6]), however, an increase in incorrect heterozygous calls can also start occurring in the 0.9-0.92 call rate range which can impact analyses that may rely on accurate heterozygous calls, such as X chromosome based sex predictions, or trio parentage tests.

Details of the call rate criteria currently used in routine AE can be found in the AE processing documentation.

12.2 X and Y chromosome sex prediction

The sex of sampled animals can be predicted using the heterozygosity of some particular SNPs on the non-pseudo autosomal region of the X chromosome, or by the existence of called SNPs from the Y chromosome. Cross-referencing these predictions against the pedigree recorded sex can help reconcile sample mismatches and assist in linking samples to the associated animals. Sex predictions can also enable quick segregation of male and female samples for use as potential sires and dams in parent discovery.

Note the "genotype sex" is an estimate based only on the sex SNP information that is available for a particular genotype and is neither guaranteed nor expected to be 100% concordant with the sex determined by inspection of an animal. While rare, it is possible for the estimated genotype gender to conflict with the actual gender of the animal. There are a variety of reasons this can occur (in no particular order):

- Mismatched/mislabeled or otherwise unreliable genotype samples.
- An inbred female can in particular cases inherit largely the same X chromosome from both of her parents. This can lead to her genotype "looking" male (i.e. homozygous sex SNPs) and being incorrectly flagged as such.
- Intersex animals (e.g., XXY, Klinefelter).
- Chimeric animals (e.g., freemartins).
- Some other sex chromosome genetic abnormality.

Assuming the sum of the above atypical cases may occur as little as 0.5% of the time, this could still lead to 5 genotypes out of every 1,000 being misclassified. As such, if the estimated genotype gender appears to conflict with the recorded sex of the animal, this suggests an investigation is only needed to determine if the genotype sample is incorrect, the recorded sex is incorrect, or the animal has a rare genotype causing it to be misclassified. Such an investigation may include checking to see if that animal has been recorded as a parent in the pedigree (and how many times) and, if necessary, contacting the breeder regarding potential infertility or abnormalities and/or resampling the animal.

Genotypes that do not have either the necessary X or Y SNPs available (e.g. low-density parentage only panels) cannot have their sex estimated.

12.3 Similarity (concordance of duplicated samples)

Multiple samples with the same animal identifier are automatically tested for concordance. Samples with <98% concordance are automatically flagged and require manual intervention to reconcile only the matching samples

[6] Cooper T.A., Wiggans G.R. and VanRaden P.M. 2013: Short communication: Relationship of call rate and accuracy of singe nucleotide polymorphism genotypes in dairy cattle. Journal of Dairy Science Volume 96, Issue 5 pages 3336-3339.

to the animal. Duplicated samples with more than 98% concordance are selected from highest call rate to lowest and merged to produce a reconciled sample for downstream analyses.

Appendix 1: Progeny Test Schemes (informative)

Before the advent of genomic selection, artificial insemination companies relied on progeny testing to identify bulls with high genetic merit. The progeny test scheme relies on phenotypic assessment of an individual's offspring to make decisions regarding genetic selection.

In a progeny testing scheme young bulls with the highest genetic merit are selected, using the average estimated breeding value of their sire and dam (commonly referred to as parent average). This has an accuracy (reliability) of only 30 to 40%.

A group of elite cows are identified as potential dams of young bulls (i.e. bull mothers). Progeny testing is necessary because most traits of economic importance in dairy cattle (e.g., milk production) are sex-limited and can be measured only in females. These bull mothers are mated to elite progeny-tested sires from the previous generation for the specific purpose of producing bull calves.

Once these young bulls reach sexual maturity, which typically occurs at about 12 months of age, they are mated to a large number of cows on commercial farms, with the goal of producing approximately 100 daughters.

Approximately 3 years later, the daughters of these young bulls will begin lactating, and this information is used to calculate the estimated breeding value of their sires for milk production and other key traits. This typically has reliability of 75 to 85%.

At this point, the bulls are approximately 4.5 years of age, and the artificial insemination companies would decide which bulls should be culled and which bulls should be marketed to dairy farmers for the purpose of siring the next generation of replacement heifers.

Overall, progeny-testing schemes are time-consuming and costly because the artificial insemination companies have to wait many years to obtain genetic predictions with sufficient accuracy for making selection decisions, and in the meantime, hundreds of bulls are housed "in waiting" while phenotypes are measured on their daughters.

Excerpt from Genomic selection in dairy cattle: Integration of DNA testing into breeding programs (https://academic.oup.com/af/article/2/1/4/4638584).

Appendix 2: Genetic Gain and Estimating Breeding Values in a Multibreed Context (informative)

Genetic gain is accomplished by breeding genetically superior animals. Ranking animals for selection in a multibreed context requires estimating within-breed breeding values (EBVs), and estimating the breed effects which are multiplied by each animal's individual breed proportions. The resulting sum is added to an animal's within-breed EBV to generate an across-breed EBV. The accuracy of the across-breed EBV is determined by the accuracy of the within-breed EBV, the accuracy of the estimated breed effects, and the precision of the defined breed proportions.

The EBV of an animal with no recorded performance or progeny information is trivially calculated as the average of its parents' across-breed EBVs. These are calculated by adding the estimates of the parents' within-breed EBVs plus the estimate of the animal's own breed effects (computed as half of its sire's and dam's breed effects).

Accurately estimating an animal's EBV is highly dependent on recording accurate pedigree information for that animal and its ancestors. This is even more important in a multibreed context where estimating the breed proportions of an animal is dependent on the recorded pedigree of that individual, and estimating the breed effects is dependent on accurate phenotypic measures of animals who have precisely recorded breed proportions.

If the parentage of an animal is known, then the expected breed proportions of that animal can be assigned as half of its sire breed proportions plus half of its dam breed proportions. Precise recording of breed proportions occurs naturally when and animal's actual parentage is recorded, provided that the breed proportions of the population founders are precisely known.

Genetic evaluation using actual rather than presumed parentage is critical to genetic gain. Without accurately linking the progeny to its dam and sire, the parent average prediction of the progeny genetic merit will be biased. If it is biased upwards, the progeny will have a better evaluation than it deserves and may be selected as a parent when it should not have been. If it is biased downwards, the progeny will have a poorer evaluation than it should have had and this may lead to it being overlooked for selection.

Incorrect parentage recording of progeny will also bias the evaluation of the parents once the progeny has a phenotypic record.

Recording actual parentage and therefore breed is a key requirement to achieving world-leading rates of genetic gain, especially in a multibreed context.

Appendix 3: Certified Data Providers for Parentage and Breed Recording and Genetic Testing Data Standard (informative)

Relationships between Certified Data Providers for the Parentage and Breed Recording and Genetic Testing Data Standard and the data flows from sample collection to delivery for results (Figure A3.1). The Certification standard for Certified Data Provider – Nominator for Genetic Testing for Cattle to the Dairy Industry Good Animal Database ('CDP Nominator') liaises with the clients to assist with or submit samples to a certified laboratory for genetic testing. A certified laboratory is approved using the Certification standard for Certified Genotype Data Provider for Cattle to the Dairy Industry Good Animal Database ('Certified Genotype Data Provider').

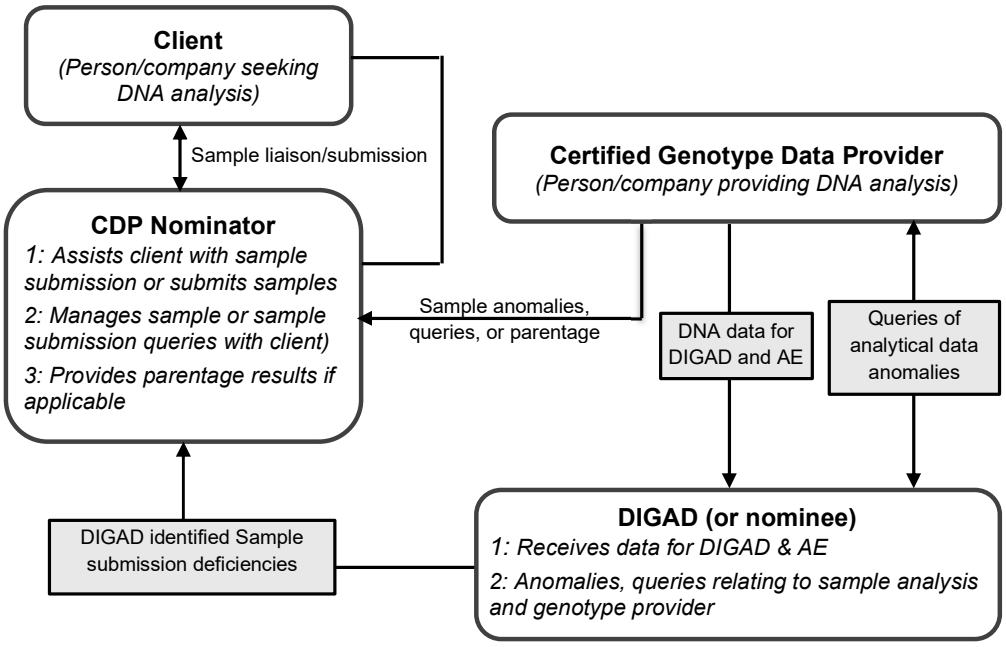

Figure A3.1 *Relationships between Certified Data Providers for the Parentage and Breed Recording and Genetic Testing Data Standard and the data flows from sample collection to delivery of results.*