ICAR Guidelines for Parentage Verification and Parentage Discovery Based on SNP Genotypes

Prepared by: ICAR DNA Working Group

Purpose

In late 2017 or early 2018, ICAR will commence offering two new services related to the use of SNP genotypes for dairy and beef cattle. One of these is the **Geno**type **Ex**change **P**arentage **S**NP **E**xchange (GenoEx-PSE) Service, which allows countries to exchange SNP genotypes for the purpose of offering parentage analysis services to specific breed populations at the national level. Another new service, which is also a prerequisite for GenoEx-PSE, is the service of ICAR Parentage Analysis Accreditation for Data Interpretation Centres, which allows organizations wishing to carry out parentage analysis services to be accredited by an independent third party. Organizations wishing to be a service user of GenoEx-PSE must first receive the ICAR Accreditation for Data Interpretation Centres.

In July 2012, the International Society for Animal Genetics (ISAG) approved standards for genotyping laboratories to conduct parentage verification. These *"Guidelines for cattle parentage verification based on SNP markers"* were based on carrying out parentage verification using a set of 200 recommended SNP, with 100 considered as core SNP and a second set of 100 as backup SNP.

The purpose of this document is two-fold:

1. Revise the existing ISAG guidelines for conducting parentage verification in dairy and beef cattle using the 200 SNP recommended by ISAG, and

2. Establish new guidelines for conducting parentage discovery based on the SNP included for this purpose in the GenoEx-PSE genotype exchange service, which initially totals 554.

Principles

When carrying out parentage analysis, which includes parentage verification and/or parentage discovery, the following underlying principles should be considered:

- A consistent set of SNP must be defined for use by all organizations for international recognition of parentage analysis accreditation and for the subsequent delivery of "certified" parentage information.

- Each animal involved in the parentage analysis process (i.e.: animal and each potential parent) must have a SNP genotype available for which a minimum proportion of the defined set of SNP have been called and are available. When establishing such minimum requirements consideration should be given to the inclusion of the defined SNP on various genotyping chips used widely in the population of animals being considered.

- While it is understood that only informative SNP (i.e.: SNP whereby the animal and the parent in question are both homozygous) provide useful information for parentage analysis, it is more practical that guidelines are based on the total number of SNP available for the animal and parent(s) in question. Roughly speaking, about one-third of available SNP for parentage analysis are informative but this proportion depends on the average minor allele frequency of the included SNP within the population of animals being considered.

Parentage Verification

The current ISAG guidelines for parentage verification in cattle are based on a set of 100 core SNP and then a second set of another 100 backup SNP. The guidelines outline a possible two-step process whereby results from the 100 core SNP may be used to either deem the status as "Parentage Accepted" or "Parentage Excluded" and only when this first status is "Parentage Doubtful" does the analysis continue using the second group of 100 backup SNP. This second step of the parentage verification process may then result in a final status of "Parentage Accepted" or "Parentage Excluded". In the event that a second result of "Parentage Doubtful" arises, and new samples and/or genotypes for the animals involved have been processed and the customer cannot identify any other possible parents, the recorded parentage would still be deemed as "Parentage Accepted".

Given the current state of knowledge and experience with using SNP genotypes for parentage verification, the following modifications to the existing ISAG guidelines have been approved for the purposes of ICAR accreditation:

1. Upon approval by the ICAR DNA Working Group, specific individual SNP included in the current group of 200 SNP recommended by ISAG for parentage verification in cattle, may be deemed inappropriate for inclusion. This reduction in the total set of SNP to be used would be applied by all organizations receiving ICAR Accreditation for Data Interpretation Centres. Annex 1 attached lists the SNP that have shown to cause problems for parent verification and are recommended for exclusion from the original list of 200 recommended by ISAG for parentage verification. On an ongoing basis ISAG may identify other problematic SNP from various chip platforms and technologies for inclusion in Annex 1 upon final approval by the ICAR DNA Working Group.

2. Given the improved accuracy of parentage verification achieved by the inclusion of more SNP, the current two-step process should be replaced by a single analysis based on the full set of approximately 200 SNP for parentage verification, excluding those in Annex 1.

3. The required minimums in terms of number of SNP, as outlined in the current ISAG guidelines, must be scaled to reflect the total number of SNP to be used for parentage verification analysis. For example, if the total number of SNP from those recommended by ISAG, is reduced from 200 to 196 (as is the current situation), then the minimum number of SNP available in the profile of each animal and potential parent must be scaled to 186 (i.e.: 95% truncated down), based on the current ISAG minimum of 95 out of 100.

4. For assigning the parentage verification status according to the number of SNP conflicts found, the revised rules approved by ICAR are the following:

   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 186 SNP available for the animal and each parent, the minimum number of common SNP available for verifying each animal-parent combination is 176 (i.e.: 196 - (2 x (196-186)) = 176).

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

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

   Step 2: In the case that both sire and dam have a status of "Parentage 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 166 (196 - (3 x (196-186)) = 166).

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 => Parentage Accepted
- Number of mismatches/SNP conflicts: 4 - 7 => Parentage Doubtful
- Number of mismatches/SNP conflicts: >7 => Parentage Excluded

5. For animals with only one parent genotyped, only those animal-parent combinations achieving the status of "Parentage Accepted" from Step 1 would qualify for the organization to issue an official confirmation of parentage for that parent. For animals with both parents genotyped, only those animals achieving the status of "Parentage Accepted" from Step 2 would qualify for the organization to issue an official confirmation of parentage including both parents.

6. As an added service for those organizations receiving ICAR accreditation to carry out parentage discovery, the process outlined below could be applied to all animals for which the parentage verification result was either "Parentage Doubtful" or "Parentage Excluded" in either Step 1 or Step 2 above.

Parentage Discovery

No international guidelines currently exist for organizations to carry out parentage discovery even though most, if not all, genetic evaluation service providers have developed such processes internally. As with parentage verification, the accuracy of parentage discovery is improved as the number of SNP included increases. For the GenoEx-PSE service, a list totalling 554 SNP has been defined for the genotype exchanges involving service users that have been accredited by ICAR for this level of parentage analysis and have agreed to upload these SNP to the GenoEx-PSE database at the Interbull Centre, which is the requirement for downloading the same. These 554 SNP include the 200 SNP recommended by ISAG for parentage verification in cattle as well as an additional group of 354 SNP. In addition to the 200 SNP for parentage verification another 75 for parentage discovery are spread across chromosomes 1 to 29 while the remaining 279 SNP were selected from only ten chromosomes, specifically 1, 2, 3, 5, 7, 8, 11, 13, 19 and 21. This strategy for SNP selection was adopted to reduce the accuracy of genotype imputation and genomic predictions in the event that a GenoEx-PSE service users attempts to use the exchanged genotype in this manner even though it is clearly prohibited as outlined in the GenoEx-PSE Service Agreement.

To be consistent with the principles and revised guidelines for parentage verification outlined above, the following are guidelines approved by ICAR for parentage discovery:

1. The ICAR DNA Working Group may, from time to time, identify and approve SNP, from among those included in the GenoEx-PSE service, that must be excluded for carrying out parentage discovery, which are listed in Annex 2 attached. Any such SNP would include those approved for exclusion for parentage verification and may also include other SNP once there is sufficient reason to do so.

2. Organizations 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 genetically identical animals are pre-identified such that a discovered parent is reported as any one of the genetically identical siblings

3. Based on a recent assessment of SNP lists associated with various SNP chips used internationally to genotype dairy and/or beef cattle, each chip has at least 500 in common with the 554 SNP recommended for parentage discovery. Given possible call rates of genotypes for the animal and any potential parent to be discovered, it has been approved by ICAR that each genotype included in such an analysis has a minimum of 450 of the 554 SNP available in order to conduct parentage discovery.

4. 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.: male older than animal with fewest conflicts) or dam (i.e.: female older than animal with 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, a minimum number of common SNP between the animal and each parent will be 350.

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

   - Percentage of common SNP with a conflict: 0 to <1.0% => Parent Discovered
   - Percentage of common SNP with a conflict: 1.0 to <3.0% => Parent Doubtful
   - Percentage of common SNP with a conflict:: ≥3.0% => Parent Excluded

   Step 2: In the case that an animal has both a sire and dam with a successful status of "Parent Discovered" from Step 1, this parent combination must also be verified.

   For this step, the following rules apply for assigning the status of the combination of parents discovered:

   - Percentage of common SNP with a conflict: 0 to <1.5% => Parents Discovered
   - Percentage of common SNP with a conflict: 1.5 to <4.0% => Parents Doubtful
   - Percentage of common SNP with a conflict:: ≥4.0% => Parents Excluded

5. For animals with only one parent with the status of "Parent Discovered" in Step 1, only that animal-parent combination would qualify for the organization to issue an official confirmation of parentage for that parent. For animals with both parents with the status of "Parent Discovered" in Step 1, only those animals achieving the status of "Parents Discovered" from Step 2 would qualify for the organization to issue an official confirmation of parentage with the status of "Parentage Accepted" for both parents.

ANNEX 1: SNP Among the ISAG 200 to be Excluded from Parentage Verification for ICAR Accreditation

SNP Name (Illumina Bead Chips)	ISAG Group	Reason for Exclusion
ARS-USMARC-Parent-DQ837645-rs29015870	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

SNP associated with other chip platforms and/or technologies may be added over time.

ANNEX 2: SNP Included in GenoEx-PSE to be Excluded from Parentage Discovery for ICAR Accreditation

SNP Name (Illumina Bead Chips)	ISAG Group	Reason for Exclusion
ARS-USMARC-Parent-DQ837645-rs29015870	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

SNP associated with other chip platforms and/or technologies may be added over time.