

Methods for discovering and validating relationships among genotyped animals

G.R. Wiggans,¹ P.M. VanRaden,² and L.R. Bacheller²

¹ Council on Dairy Cattle Breeding, Bowie, Maryland, USA

² Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland, USA



United States
Department of
Agriculture

Validation of parents

- **Over 2.2 million animals genotyped in U.S. system**
- **Portion of parents validated**
 - **97% of sires**
 - **39% of dams**
- **Each genotype compared with all others to discover identical genotypes and parent-progeny relationships**
- **Animals with incorrect sire or dam excluded from evaluation**

Validation of grandsires

- **If parent not genotyped or not confirmed, grandsire is checked**
- **Grandsire declared unlikely if animal and grandsire have more opposite homozygotes than threshold, which declines as possible comparisons increase**
- **Possible grandsires are suggested if low percentage of conflicts and birth date reasonable**
- **Animals with unlikely grandsires excluded from evaluation**

Detection of chromosomal abnormalities

- Where parent and progeny have more conflicting SNPs than allowed for a true parent-progeny relationship, location of conflicts is checked
- If conflicts are concentrated on a single chromosome, parent-progeny relationship is accepted
 - Large deletion – animal is homozygous in the region
 - Uniparental disomy – heterozygous SNPs in the region
 - 102 cases discovered so far

Quality control

- Each SNP evaluated for
 - Call rate
 - Portion heterozygous
 - Parent-progeny conflicts
- Parent-progeny conflicts assessed for all SNPs in common between parent and progeny genotypes
- Trio test if both parents genotyped
- 30 chips supported

Computational burden

- **Computer time to compare each genotype with all others steadily increases with number of genotype in database**
- **1,000 SNPs that were on all chips used to exclude most unrelated animals**
- **Further speed-up needed**
 - **Compare fewer SNPs**
 - **Exclude some genotypes from comparison**
 - **Optimize comparison method**

100 SNPs

- Selected based on call rate, MAF, and Mendelian consistency
- Measure: $\text{Conflicts} / (\text{number of both SNPs homozygous})$
- Threshold of 8.4% eliminated 99.7% of genotypes without eliminating any confirmed parent-progeny pairs
- Test with only 50 SNPs eliminated only half the unrelated animal genotypes

Compare genotypes for fewer animals

- For animals with both parents confirmed, check only recent genotypes (starting with births 500 days before) for identical genotypes
- For animals with 1 parent confirmed, skip genotypes with a different confirmed parent when checking for identical genotypes
- For grandsires, skip comparisons with bulls that have no progeny

MGS checking with haplotypes

- For animals included in the evaluation, haplotypes are generated during imputation
- These haplotypes can be used to validate or discover MGS more accurately (even MGGS can be discovered)
- For MGS, identify bulls with around 45% of haplotypes in common and at least 15% better than next best bull
- Discovered MGS assigned as dam's sire if unknown

Use haplotypes for initial MGS discovery

- **Remove searching for possible MGS from initial genotype validation program for faster processing**
- **Include new animals with unknown or unlikely MGS in weekly evaluation calculations (confirmed sire required)**
- **For genotypes not qualifying for evaluation, blank conflicting pedigree and suppress release of evaluation**
- **Continue use of current SNP comparison process for PGS**

Timing comparison

- Time to load 1 submission of 1,967 genotypes
 - Current – **51 minutes**
 - Eliminate 497 MGS searches – **39 minutes**
- Time to run weekly MGS discovery for Holsteins – **9 minutes**
- Time to run monthly MGS/MGGS discovery for Holsteins – **7 hours**

Further possible use of discovered MGS

- When dam is unknown, constructed ID necessary to store discovered MGS

Ayrshire	Brown Swiss	Guernsey	Holstein	Jersey
21	245	68	213,704	21,963

- More complete pedigree gives better imputation
- Numerator relationship matrix (A) more similar to genomic relationship matrix (G)

Conclusions

- **Rapid increase in size of genotype database requires periodic modification of procedures**
- **Checking all genotypes is desirable for correctly assigning animal to genotype and improving pedigree accuracy**
- **100 high quality SNPs are effective in excluding most genotypes that are not parents or progeny**
- **Grandsires (even great-grandsires) can be checked and candidates discovered**

Acknowledgments & disclaimers

- **USDA-ARS project 8042-31000-002-00, “Improving dairy animals by increasing accuracy of genomic prediction, evaluating new traits, and redefining selection goals”**
- **Mention of trade names or commercial products in this presentation is solely for the purpose of providing specific information and does not imply recommendation or endorsement by USDA; USDA is an equal opportunity provider and employer**



Questions?