Maternal grandsire verification and detection without imputation

Jan-Thijs van Kaam - Anafi
Ben Hayes - Department of Primary Industries, Dairy Futures CRC
Pedigree reconstruction with SNP

- For a genotyped individual, sire and dam can be “verified or detected” very accurately even with LD genotypes.

- In dairy cattle, dams are often not genotyped.

- Can we detect maternal grand sires (MGS) if dams are not genotyped?

- Here we propose a haplotype based approach.
Duo & Trio for MGS checks
(VanRaden et al. 2013)

- **Duo**: Compare MGS genotype with animal genotype.
- **Trio**: Compare MGS genotype with animal genotype, while for heterozygous loci, using sire genotype to determine which allele the animal received from its dam.
  - Less conflicts ➔ better candidate MGS.
  - Different threshold levels for Trio and Duo and for different SNP chips needed.
  - Trio more powerful than Duo but less than haplotypes ➔ additional information in lengths of shared haplotype.

24/8/2013
Interbull meeting Nantes
Haplotypes for MGS checks

More power needed... Let’s use haplotypes:
- Joint inheritance (IBD) of multiple markers
- More markers ➔ More powerful

However animals are genotyped on SNP chips with different densities... Let’s impute.
Imputed haplotypes for MGS checks

- Using phasing and imputation to obtain haplotypes increases computational time a lot.

- Disrupts workflow:
  1. Receive genotype data
  2. Optionally duo/trio test $\rightarrow$ MGS (initial test)
  3. Phasing/Imputation
     - How often/month?
  4. Haplotype test $\rightarrow$ MGS (final test)

- Initial and final test result might differ.
Haplotype based MGS checks without phasing and imputation

1. For target individual, derive dam haplotype
   Sire genotype
   AA  AT  CC  GG  AC
   Individual genotype
   AA  AT  CT  GC  AA
   Individual’s haplotype from dam
   A  ?  T  C  A

2. For each possible MGS, how many haplotype matches of x markers?
   Possible MGS genotypes compared to individual’s dam
   AA  –  AT  CG  AA
   AA  –  AA  CG  AA

24/8/2013  Interbull meeting Nantes
Haplotype-based MGS check test statistic

Decisions for computing a test statistic:
1. How many haplotypes should match across the genome?
2. What should the length of the segments be (number of SNP)?

- Chips of different density converted to same chip. Missing SNP genotypes give no conflict.
Gamete creation

- **No crossover:**
  - Parental gametes

- **Crossover:**
  - Recombinant gametes
Crossovers

- Maternal grandsire ➔ Dam ➔ Animal
  - 3 generations
  - 2 transmissions
  - 2 meioses
  - Double crossovers have low frequency
  - Few crossovers between MGS and animal
  - Very long segments can be used
Number of matching haplotype segments

Segments of half an autosome:

- Simulation of crossovers during gamete creation shows that grandparents have 20.0 matching segments.
- Actual data showed that pedigree MGS on average had 20.8 matching segments.
- MGS with ≤ 14 matches are considered in doubt and with ≤ 6 as likely wrong.
Haplotype based MGS checks without phasing and imputation

Alternative approach:
- Instead of more markers in same segment obtained by imputation, use longer segments without phasing and imputation
  - No phasing → saves time
  - No imputation → saves time
  - No imputation errors
  - Single test, no initial and final test
  - Works across SNP chips
    - No need for different thresholds
- More chance of recombinations

24/8/2013
Interbull meeting Nantes
Maternal grandsire verification and detection: 2 phases

1. Verification: Confront animals with pedigree MGS to select the animals with little haplotype similarity or without genotyped pedigree MGS

2. Detection: Create an ordered list of candidate MGS for the selected animals

In each step a test statistic is computed for each combination of interest
Selection of suitable candidate MGS uses some rules

- Only males can be MGS
- Animals cannot be their own ancestor
- Current maternal granddam's maternal sibs are excluded as candidate MGS
- Current maternal granddam's sons are excluded as candidate MGS
- Minimum 1-generation interval > 600 days between animal and dam and between dam and MGS
- Minimum 2-generation interval > 1200 days in case dam’s birthdate is unknown
Some results

- 4000 genotyped animals (74% LD genotypes) born in 2013 with sire and MGS genotypes and without dam genotype and after cleanup
- Propose maximum 4 candidate MGS from on average 24000 and compare with pedigree MGS

<table>
<thead>
<tr>
<th>Segment length IBD between Animal &amp; MGS</th>
<th>Pedigree MGS as first candidate MGS</th>
<th>Pedigree MGS in 4 candidate MGS list</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50 chrom. ~ 715 SNPs</td>
<td>98.22%</td>
<td>99.97%</td>
</tr>
<tr>
<td>0.33 chrom. ~ 477 SNPs</td>
<td>98.13%</td>
<td>99.65%</td>
</tr>
<tr>
<td>0.25 chrom. ~ 357 SNPs</td>
<td>98.13%</td>
<td>99.57%</td>
</tr>
<tr>
<td>0.10 chrom. ~ 143 SNPs</td>
<td>97.90%</td>
<td>99.47%</td>
</tr>
<tr>
<td>500 SNPs</td>
<td>97.88%</td>
<td>99.55%</td>
</tr>
<tr>
<td>75 SNPs</td>
<td>97.40%</td>
<td>99.25%</td>
</tr>
<tr>
<td>Trio</td>
<td>82.18%</td>
<td>87.50%</td>
</tr>
</tbody>
</table>
SNP chips enable MGS verification and detection

Very long haplotypes can do a good job without phasing and imputation

Thank you for your attention!

Questions?