INTERBULL TECHNICAL WORKSHOP

ESTABLISHING THE FRAMEWORK FOR INTERNATIONAL GENOMIC EVALUATIONS

Guelph, Ontario, Canada
February 27 and 28, 2011

EXECUTIVE SUMMARY

A total of 84 representatives from 25 countries gathered at the Delta Hotel and Conference Center, in Guelph, Canada, to participate in the Interbull technical workshop. The program was structured to have an initial warm-up session, where the members of the Scientific Advisory Committee presented their views for the future of international genetic evaluations in the era of genomics. The next session consisted of reports on the latest developments on the Intergenomics project, the GEBV test and GMACE. Then the participants were divided into six discussion groups that addressed important questions surrounding genomic evaluations: validation of GEBVs, genomic reliabilities, imputation and the various SNP panels, use of MACE results of input for genomic evaluations and consequences of moving away from progeny testing. In the second day an important panel took place to discuss different experiences and views about establishing common repositories of genotypic data and the respective role of Interbull. Last, but not least, presentations on future developments for Interbull and also a few spontaneous presentations ended the event. As usual in Interbull events, both the Interbull Technical Committee (ITC) and the Interbull Steering Committee (SC) had their ordinary meetings after the workshop. This summary is divided in two parts: the most important decisions made by the SC and the detailed highlights of the workshop.

GENERAL NEED FOR INTERNATIONAL GENOMIC COMPARISONS

Countries participating in Interbull evaluations can be classified in three groups:

1. Countries with their own bull selection schemes and genomic evaluation system
2. Semen importing countries with their own genomic evaluation system
3. Semen importing countries without a genomic evaluation system

A natural consequence of the above is to have different needs for international comparisons, which may require different solutions and result in different service demands from Interbull (such hypothesis was already established during the Interbull Technical Workshop in Uppsala, January 2009). For instance, data sharing and exchange of genotypes may suit the needs of groups 1 and 2, but do not resolve the problem for group 3.

Therefore an implementation plan for solving these needs was developed and the Steering Committee decided on the necessary steps to have services in place.

MAIN STEERING COMMITTEE DECISIONS

- Interbull Technical Committee composition
  - Annual meeting in Stavanger: end of Hans Wilmilk mandate in the SC
  - Sucessor for ITC chairman: Gert Petersen Aaman

    - Two new members: Gerben de Jong and Peter Sullivan
- S-GMACE Pilot run
    - March 2011 – new set of clear instructions to countries on data preparation
    - April 1, 2011 – new data call for S-GMACE pilot
    - May 10, 2011 – data reception deadline for the S-GMACE pilot
    - June 15, 2011 – S-GMACE distributed to participating countries
    - August 26-28, 2011 (Stavanger)
        - Presentation of results in the Interbull open meeting
        - ITC + invited experts from participating countries evaluate results and make recommendation to the SC
- Intergenomics
    - ITBC can provide genomic evaluations as a service to the BSW populations (7 votes to 1)
    - March 2011 Pilot
        - Comprehensive report must be sent to ITC for final technical recommendation to SC
- GEBV test
    - Official Calendar 2011 for protein yield
        - May 9, 2011 (data reception); June 6, 2011 (official release)
        - July 4, 2011 (data reception); August 1, 2011 (official release)
        - October 31, 2011 (data reception); November 28, (official release)
    - New set of rules
        - New validation needed
            - Major changes in natiional evaluation
            - Major change in reference population
            - Every two years
        - Significance level
            - Statistical significance (±2 SE), OR
            - Biological significance (Confidence Interval: ±0.1)
    - Countries will be stimulated to submit validation data for milk and fat yield and key traits other than production to help further evaluation of the method (sta, scs, dlo, cc1)
        - SC will decide if validation for EU matters must be expanded to Milk and Fat yields (Stavanger, Norway)
        - SC will decide how results for validation of other key traits will be used
- Common repository of genomic information at Interbull
    - Concept of establishing a data base is approved, but content and rules need to be discussed and decided after a feasibility plan of ITBC has been presented
    - ITBC must present a feasibility plan/business plan for such data base
        - National experts must be consulted for technical specifications
        - IGenoP proposed settings should be considered
        - Include a proposal for parentage verification
- ITBC must launch survey on genotyping and sequencing
- Sire-Dam pedigree in MACE
    - In principle, should be implemented in the April 2011 routine run
    - ITBC must provide explanation for some changes observed by NGECs until March 12, 2011
    - ITBC must write an extension article explaining the modifications made in the MACE evaluations and the consequences of adopting a more complete pedigree

DETAILED HIGHLIGHTS FROM THE WORKSHOP

VIEWS FROM THE SCIENTIFIC ADVISORY COMMITTEE

- Significant changes in standards in the near future:
    - Shift from 50K to HD SNP panel(s)
    - Shift from BLUP methods to Bayesian methods, due to the size of the reference populations and the number of SNPs
    - Shift from single breed to multi-breed SNP prediction
    - Shift from genetic evaluation of individuals (EBVs) to evaluation of individual SNP effects
    - Further, SNPs will be replaced by actual QTLs and their alleles in the near future through complete genome sequencing
    - Concepts about inbreeding will need to be revised, and the models for genetic evaluation of individuals may become more biased due to using inappropriate relationship matrices
- Urgent need for collaboration:
    - Combination of genotypes and sequencing data from different sources, less potential of exchange of raw phenotypic data
    - Opportunity to invest on a common pipeline for data processing (methods, programs)
    - Data from HD SNP panels and sequencing must be used to investigate associations with biological information, based on the human genetics experiences
- Methods for genetic evaluation:
    - The use of the A matrix in genetic evaluation needs to be reconsidered, because of the changes about the concept of inbreeding and because sons will no longer be random samples of their sire
    - Instead of evaluating individuals, national genetic evaluations should be more about estimating the effects of alleles of all QTLs, and their interactions
    - No universal method of estimation is likely to appear, and there is a need to consider different models and methods per breed, trait, population
    - Neural networks have been shown to have a great potential for prediction models of complex traits.
    - Evaluations must consider the bias due to pre-selection of candidates.
- Interbull role:
    - Establish the network between groups that are sequencing bovines and may be interested in collaboration
    - Monitor genetic diversity in the dairy cattle populations to avoid loss of alleles, if access to genotypes is granted
    - Facilitate exchanges of genotypes, GEBVs, pedigrees, genetic diversity measures, lists of genotyped/sequenced individuals
    - Ensure correct international genetic merit estimates for reference populations and marketed bulls
    - Subcontract genomic evaluations of smaller breeds

PROGRESS REPORTS ON GENOMIC INTERNATIONAL COMPARISONS

GEBV TEST

- Most applicants passed the official test for protein yield (August 2010 and November 2010 tests)
- R^2 values

    - Rather low, especially with small test bull sets
    - $R^2_{\text{Model 1}} > R^2_{\text{Model 2}}$, as expected
- Works for different traits
- Selective genotyping
    - Use of selection intensity to predict $E(b_1)$ seems to be consistent

SIMPLIFIED GENOMIC MACE (S-GMACE)

- Call for data (December 2010)
    - HOL populations passing the GEBV test for protein
    - Traits: pro, sta, scs, dlo, cc1
- Implementation:
    - The procedure works using regular MACE machinery
    - Major difficulty: harmonization of input data in order to fairly select among GEBVs from several sources
    - High correlations with conventional MACE EBVs
    - Plausible number of young bulls with no daughters appear on top-100 lists
    - Increase in average international reliability when comparing run including GEBVs to run not including GEBVs
- Overall conclusion:
    - Further investigation of potential biases
    - Promising option while GMACE is not implemented (short term)

GMACE

- Simulation results
    - Encouraging first results were from pretty “ideal” simulated data, better represents future than current situation
- Implementation issues
    - Not all animals are genotyped (selective?)
    - Residual correlations extend from animal to familyGEBV definitions vary among countries
    - MACE data are used for SNP training
    - Accounting for correlated residuals still not fully solved
- Overall conclusion: more research is needed

INTERGENOMICS

- Pilot results
    - Pooling of genotype data from several countries is an effective way of building a large enough reference population
    - International (MACE) breeding values work well as the input/phenotype
    - Adding genomic information improved reliability of estimates for all trais and ages.
    - Estimated reliability gains were overestimated when compared with gains obtained from validation models
- Business plan proposed by BSW populations to Interbull
    - Service description
        - Prediction equations for all traits available at Interbull for the BSW breed, based on all genotypes available (3 times/yr, following MACE routine runs)
        - Validation system coherent with the official international checks

    - Time table
        - Last pilot: March 2011
        - First run: April 2011
        - Official considered official only after passing the GEBV test
    - Service fees
        - Separate from regular fees
        - Partners and ITBC already agreed in a model for sharing costs

DISCUSSION GROUPS

GENOMIC VALIDATION OF NATIONAL SYSTEMS

- Interbull guideline to help countries find best balance between size of reference versus validation groups
- 4-year cut-off a challenge for:
    - New traits
    - Small populations
- Large variation in predictive ability of genomic evaluation systems
    - Statistical methods
    - Size of reference population
- Develop a strategy to communicate better/in a simpler way with the industry

GEBV TEST FOR ALL TRAITS

- 427/2006 EU decision: Responsibility of ICAR / Interbull
    - Validate also milk and fat yield
- No reasons to handle it differently to MACE once S-GMACE is implemented
    - Test all individual traits and selected conformation traits
    - Every time changes in model occur or every 2 years
    - Acceptance level: combine statistical and biological significance

GENOMIC RELIABILITIES

- Definition
    - Amount of information used to estimate DGV
    - Reliability should reflect predictive ability of the model
- Theoretical reliabilities should be adjusted by using results from cross validation
- Cross validation reliabilities for young bulls must be higher than PA reliabilities
- Harmonization
    - No standard recommendation of how to combine conventional and genomic reliabilities yet

SNP PANELS / IMPUTATION

- Promote collaboration
    - Among research groups dealing with sequence data
    - Imputation validation data sets could be shared to evaluate results
    - Sharing of SNP map corrections
    - Sharing Affimetrix genotypes would speed up imputation between Affimetrix and Illumina panels
- Interbull role

  - Provide guidelines on SNP quality (library of SNP corrections)
  - Provide standards for data exchange formats
  - Provide programs that facilitate conversion between SNP panel versions
- SNP for parentage verification
  - Pressure from industry to move away from ISAG microsatellites to SNP (~120)
  - Clear demand for ICAR / Interbull to establish a standard if ISAG is not acting fast

USE OF MACE RESULTS AS INPUT FOR GENOMIC MODELS

- Stimulate exchange of methods for usage of MACE EBVs as input for genomic predictions
- Additional information from MACE runs
- Recommendations should come from ITC

MOVING AWAY FROM PROGENY TEST SCHEMES: CONSEQUENCES ON CONVENTIONAL (INTER)NATIONAL EVALUATIONS

- Discontinued progeny test means “unorganized” progeny test (decrease in incentives)
- Availability of phenotypic records
  - Phenotypes collected for
    - Management purposes: little impact
    - Mainly calculation of conventional EBVs (eg. type): difficult to maintain
  - New traits
    - More interest due to new opportunities with genomics
    - Need for harmonisation of trait definition/data collection
  - Impact: bias due to pre-selection
- Interbull role
  - Pedigree / genotype repository
  - Good quality (bias “corrected”) data for international (male and female) reference populations
  - Standard procedures and rules for fair comparisons
  - Coordination of new phenotypes recording
  - Tools for management of genetic variability

PROPOSALS OF FUTURE DEVELOPMENTS

INTERNATIONAL GENOMIC EBVS - LONG TERM PLAN (MIKE GODDARD)

- In the long term most animals will have full genome sequence
- Relationship based EBVs less accurate and just as costly as SNP based analyses and don’t capture across breed information well
- SNP based analyses are more accurate, capture across breed information, add to and benefit from biological knowledge
- Interbull could coordinate a SNP based international analysis
- Exchange SNP solutions instead of animal solutions

PANEL: COMMON REPOSITORY OF GENOTYPES AT INTERBULL

- IgenoP

  - Establishment of a database of shared genotypes at the Interbull Centre
  - Proposal has been defined in details by proponents and discussed with potential users
  - Prototype established in Ireland
  - Interested NGECs could have a service available at ITBC quickly
- Eurogenomics
  - Interbull role
    - Organizing workshops for scientific discussion
    - List of ID of young bulls
    - Encouraging exchanges of genotypes if the importing countries allows it
    - Work on general comparison of GEBVs across countries
- North American Partners
  - Interbull role
    - List of genotyped animals
    - Storage of genotypes for parent verification (120 SNPs)
    - Parent verification service
    - International comaprisons of GEBVs (GMACE)
    - Storage of genotypes and imputation services from breeds other than Holstein
- Intergenomics
  - Common genotype data base
  - ITBC provides predictions in all scales
  - Future possibilities
    - Participants have access to genotypes
    - Customized data checks
  - Clear option for small populations
- Overall discussion
  - 50K SNP panel data
    - List of animals with genotypes: all in favour
    - Genotypes of marketed bulls for use in importing countries: no problem
    - Need for establishing clear publication rules that all countries accept to follow Exchange of genotypes of reference bulls
      - No consensus, especially for large Holstein populations
  - 3K / 120 SNP panels data
    - List of animals with genotypes: all in favour
    - Interbull as a clearing house for breed associations needs: possible
    - Parentage verification services: possible
  - HD / full sequence data
    - More openness to exchange data
    - Interbull will launch a survey to evaluate the current status and promote interaction between research groups