Incorporation of genomic selection into the Murrah buffalo breeding program
Incorporation of genomic selection into the Murrah buffalo breeding program

Karen Marshall¹, Raphael Mrode¹, Poonam Sikka², Kanwar P Singh², Inderjeet Singh² and Anurag Bhardwaj²

1 The International Livestock Research Institute, Nairobi, Kenya
2 The Central Institute for Research on Buffalos, Hisar, India

December 2017
## Contents

1. Purpose 1

2. Overview of the current Murrah buffalo breeding program 2

3. Recommendations 4
   3.1. Data digitisation 4
   3.2. Best Linear Unbiased Prediction 4
   3.3. Selection index 5
   3.4. Optimal contribution theory 5
   3.6. Genomic selection 6
   3.7. Reduced genotyping array 7
   3.8. Restructure breeding program to better capitalise on genomic selection 8
   3.9. Continual review and refinement of the breeding program. 8

4. Concluding comments 9

Annex 1. Additional details of the ICAR-CIRB led Murrah buffalo breeding program 10
1. Purpose

This document provides recommendations for the incorporation of genomic selection into the Murrah buffalo breeding program—led by the Indian Council of Agricultural Research (ICAR) and the Central Institute for Research on Buffaloes (CIRB). It is the output of the collaborative project between ICAR, CIRB and the International Livestock Research Institute (ILRI): ‘Genomic techniques to profile and improve productivity and resilience in buffalo’. This report is intended to be a ‘living document’ in that the recommendations may be refined / altered over time as new information becomes available.
2. Overview of the current Murrah buffalo breeding program

The ICAR-CIRB-led program aims to produce Murrah buffalo that are genetically improved for milk yield and reproduction whilst maintaining body confirmation / type true to the Murrah breed. The breeding program operates as a dispersed open nucleus, with animals located in six institutional herds at various locations throughout the country, as well as at the field progeny testing centres in three locations. Across all institutional herds, the number of breedable females is presently 1,170. At field centres, approximate 15,000 inseminations are made annually.

The breeding program utilizes a progeny testing scheme, with progeny testing occurring both within the institutional herds and at several field sites. Bull-sets for progeny testing are selected at 18 month intervals, with progeny-test data currently available on the eleventh bull set and bull selections underway for the eighteenth bull-set.

Bulls entering the progeny test scheme are selected on a number of criteria including genetic merit of their sire based on contemporary comparison (using the daughters, and their contemporaries, 305 day first lactation yield), mother’s milk yield and reproductive performance, physical conformation to Murrah characteristics, breeding soundness and semen quality, amongst other. On occasion, outside animals—from smallholders, commercial farms, or government farms—are included in the progeny test set.

Progeny test bulls are each mated to approximately 50 females from the institutional herds, and 1,000 females belonging to smallholders in the field progeny test sites. Their daughters are subsequently recorded for first lactation yield: these typically number, per test sire, 8–10 from the institutional herds, and 6–9 from the field progeny test, thus 14–19 in total. Sires are subsequently ranked on their daughters’ milk yield, via using contemporary daughters per cent superiority [up to tenth set] and of the Harvey model [used for eleventh set], which accounts for limited relationships between the progeny test sire.

Within the breeding nucleus itself, and since 2005 nominated mating is employed where the top two–three bulls (based on the progeny test ranking) are mated to the top 25–30% elite females with best [standard lactation milk yield (SLMY) more than 2,500 kg].

The data recording scheme is comprehensive, and similar across the various institutional herds. Records comprise: weekly milk yields (until drying); milk constituents; demographic events, such as artificial inseminations, births, deaths and sales; animal weights at various ages; health treatments; disease test results; semen quality; and group feeding regimes. Records on animals in the field progeny test scheme are kept on artificial insemination and monthly milk yields (for first lactations only). Currently the majority of records on the institutional herds are kept in written herd books / registers: there is a plan for this to be digitalized into a common database under the coordination of the network improvement program. The majority of field records are in electronic format, though not in a database
The improved buffalo genetics are disseminated from the nucleus as both semen for artificial insemination and breeding bulls. Semen is sold to farmers and inseminators, the top live bulls sold to semen freezing centres across the country, and other live bulls sold to farmers for natural mating.

The key strengths of the breeding program are its:

• overall design as a dispersed open nucleus, with use of a progeny testing scheme;

• comprehensive data recording scheme and consistency in record keeping across the different institutional herds; and

• dissemination strategy for the improved buffalo genetics, including the sale of both semen and live animals.

The key weaknesses of the breeding program are that:

• data is not in a database, and thus not utilized to its full capacity;

• estimated breeding values are either not calculated (young males entering the progeny test, females) or not calculated using best linear unbiased prediction (progeny tested males), and thus rates of genetic gain are likely lower than what could be achieved; and

• optimal contribution theory is not used to maximize rates of genetic gain whilst controlling rates of inbreeding.

Full details of the breeding program are provided in Annex 1.
3. Recommendations

Recommendations for incorporation of genomic selection into the breeding program are given below. These recommendations do not need to be implemented one after the other, rather the implementation timeline can (and should) overlap.

3.1. Data digitisation

Digitization of data across all institutional herds and field progeny testing programs into a central database, that can be accessed remotely by the various centres and other stakeholders, as required. This could either comprise a customized database—such as one written in SQL which is a programming language designed for managing data held in a relational database management system—or one packaged with herd management software. The latter (herd management software) has the added advantage that it will facilitate herd management decisions as well as easily provide a variety of reports. Steps to this process are:

a. Digitization of primary data into excel spreadsheets as per the currently used registers (one excel sheet per register).
b. Checks on the primary data, including on pedigree (using software such as pedigree viewer) and the various demographic and trait data (via looking at descriptive statistics such as ranges etc.)
c. Calculation via scripting (rather than manually) of secondary or derived data and checks on these
d. Upload of the data into the customized or herd management database and checks on the upload
e. Develop and document procedures for continued use of the database, and train staff from the various centres and other stakeholders on these
f. Continued use of the database

3.2 Best Linear Unbiased Prediction

Best linear unbiased prediction (BLUP) uses all available pedigree and performance data to predict estimated breeding values (for all animals, not just progeny tested sires) and is strongly recommended as a starting point to implementation of genomic selection (which will then bring genomic information into the calculation of the breeding values). Steps to implementation of BLUP are:

a. Choice of BLUP software to be utilized—this software should be: free or low cost; easy to operate; produce accuracies / reliabilities on the EBVs; and be amenable to incorporation of genomic selection. Options include F90, Wombat, ASReml.
b. Determine the various component estimate that will be used in the BLUP analysis (from available analysis and/or literature).

c. Training of staff on use of the software in relation to predicting breeding values via multivariate BLUP, as well as generating other related data such as accuracies/reliabilities on the breeding values and inbreeding levels. This should include runs using the full breeding program data: these outputs can be used to examine breeding program progress over time (such as rates of genetic gain, inbreeding etc.).

d. Continued use of BLUP software* such that indexes based on multivariate BLUP EBVs (see below) are used to underpin all selection decisions including the selection of young bulls entering the progeny testing scheme, ranking of progeny tested bulls, and selection decisions on females.

*Note that this software will be modified on implementation of genomic selection

3.3 Selection index

A selection index combines information on EBVs for different traits (estimated via multi-variate BLUP) into a single value, such that this single value (index) can be then used to underpin selection decisions. In calculating the index the EBVs are typically weighted, with this weight accounting for the importance of the trait (such as its economic value). Use of a selection index is strongly recommended, given that the current breeding objective is multi-trait, comprising milk-yield, reproduction, and trueness to Murrah breed-type. Steps to implementation of selection index use are:

a. Based on the breeding objective, determine which traits should contribute to the selection index, and determine appropriate weights on EBVs for each selection index trait. In the first instance these weights can be based on expert knowledge: later they should be based on trait economic values determined using a profit model.

b. Generate selection indexes from the multivariate BLUP EBVs

c. Continued use of the developed selection index to underpin selection decisions (see above)

3.4 Optimal contribution theory

Optimal contribution theory uses relatedness and performance (e.g. EBV or index) information on male and female selection candidates (i.e. animals available to be used as parents) to ensure that long-term response is maximized (as mating decisions made in the current generation influence not only immediate genetic response but also long-term inbreeding). This is achieved through use of an objective function that weights genetic response against future contributions to inbreeding. In comparison to BLUP truncation selection, the use of optimal contribution theory can result in higher rates of genetic gain for the same level of inbreeding, and conversely lower inbreeding for the same level of genetic gain. Use of optimal contribution theory for selection and mating decisions is thus recommended. Steps to implementation of optimal contribution theory are:

a. Choice of optimal contribution software to be utilized — this software should be: free or low cost; easy to operate; and flexible in terms of mating options, including an option for assortative/nominated mating.

b. Training of staff on use of the software for generation of mating lists

c. Generation of mating lists

d. Continued use of the generated mating lists for actual matings.
3.5 Reference population

For implementation of genomic selection, a reference population of genotyped and phenotyped animals is required, such that the single nucleotide polymorphisms (SNP) effects can be estimated. Here the aim is to genotype at least 3,000 animals that have phenotypic information and which are linked to the breeding program, using the currently available Axiom® Buffalo Genotyping Array.

This array has 90,000 SNPs developed from eight river buffalo breeds (Italian Mediterranean, Murrah, Nili-Ravi, Jaffarabadi, Kundhi, Aza-Khel, Egyptian and Swamp type from Philippines) and subsequently verified on the Italian Mediterranean and Brazilian Murrah breeds. On the verification population 67,330 SNPs (of the total 89,988) were polymorphic (Lamartino et al. 20131). It is expected that this array will be sufficiently informative for the Indian Murrah buffalo, though this will be confirmed. Genotyping of the array will be performed by outsourcing. Steps here are:

a. Equip the existing ICAR-CIRB server to store genotype information, by installing a linux platform.

b. Determine if the currently available Axiom® Buffalo Genotyping Array is sufficiently informative for the Indian Murrah population by genotyping an initial set of 96 progeny tested bulls and subsequent analysis of the number of polymorphic SNPs.

c. Presuming the array is sufficiently informative, genotype the remainder of the reference population. As a starting point, this will comprise:

- 160 progeny tested bulls;
- 969 standing females across the centers, each with at-least one lactation record
- 200 historic females, each with at-least one lactation record
- 300 females from the field progeny test, each with at-least one lactation record
- Potentially 600 animals from a private farm, dependent on the extent of linkage of these animals to the breeding herd (tested from pedigree analysis)
- Other as they become available or are identified

d. Quality checks will be performed on the genomic data as it comes in, as follows:

- Check for minor allele frequency and determine appropriate limit
- Check call-rate of SNPs GC scores & proportion of animals with informative SNPs
- Remove SNPs not in Hardy-Weinberg equilibrium or that are on the sex chromosome

e. As genotype data builds up, it will be tested to determine if it is sufficient for genomic selection. From here a decision can be made on whether additional animals (above the initially targeted 2500) need to be genotyped.

3.6. Genomic selection

The basic assumption underlying genomic selection is that the use of single nucleotide polymorphisms (SNPs) as markers enables all quantitative trait loci (QTL) in the genome to be traced through the tracing of chromosome segments defined by adjacent SNPs. It is assumed that the effects of the chromosomes segments will be the same across the population as a result of the linkage disequilibrium (LD) between the SNPs and QTL. Thus it is important

---

1 Lamartino, D., Williams, J. L., Sonstegard, T., Reecy, J., Tassell, C. V., Nicolazzi, E. L., ... & Coletta, A. (2013). The buffalo genome and the application of genomics in animal management and improvement. Buffalo Bulletin, 32(Special Issue 1), 151-158
that marker density is high enough to ensure that all QTL are in LD with a marker. In practice, the implementation of genomic selection involves estimating the SNP effects in a reference population which consist of individuals with phenotypic records and genotypes. This is then followed by prediction of genomic estimated breeding values (GEBV) for selection candidates which do not yet have phenotypes of their own but have genotypes.

The main advantages of genomic selection include the reduction of the generation interval as young animals can be genotyped early in life and their GEBV computed for the purposes of selection. For the Buffalo situation, GEBV computed early in life can be used to select young bulls thereby reducing the cost of progeny testing. In addition, higher accuracy of GEBV, at about 20-30% above that from a parent average, has been reported for young bulls. This increases the accuracy of selection and thus rates of genetic gain.

Steps to implementation of genomic selection are:

a. Genotyping of bulls with EBVs and females with yield records with the 90K SNP chip
b. Undertake QA of SNPs for the genotyped population and select informative SNPs for genomic selection
c. Given the genotyped population and information available, define a reference population and a validation set
d. Determine the best genomic model that will suite the data and the traits to be analysed
e. It is assumed that traits to be analysed have had genetic parameters estimated otherwise these have to be estimated
f. Determine suitable software for the genomic analysis. Note that some free software such as GBLUPF90 have limits to size of data set that can be put through
g. Undertake initial genomic prediction and determine the accuracy of genomic prediction.
h. From these results determine the way forward in terms of routine genomic prediction

3.7 Reduced genotyping array

A reduced genotyping array can be developed by two approaches. Any available low density chip for Buffalo in the market can be used to genotyped subsequent animals and their genotypes can be imputed to the 90K chip. The efficiency of such imputation can be tested using available genotyped animals. The alternative is to select a reduced number of SNPs based on the results (SNP effects and accuracies) from the breeding program talking into account SNP frequencies, chromosome position and regions of the genome associated with any special characteristics of the Murrah Buffalo. This, however, requires a significant amount of analysis. Again such a chip can be imputed to the 90K, if this increases accuracy of genomic prediction. Steps to generating a reduced genotyping array are:

a. Determine if there is already a reduced SNP array in the market for Buffalo. If yes, get SNP map and determine the accuracy of imputation to 90K. If accuracy is good, then either adopt this reduced SNP for subsequent genotyping in the Murrah breed population
b. If imputation in step above is inaccurate, then undertake selection of reduced SNP for genomic prediction through various methods for selection and validation of SNPs
c. Undertake GWAS and identify any specific genomic regions of the Murrah genome associated with any special features of the breed
d. From the 2 steps above, formulate a reduced SNP set
e. Note that this option involving the last 3 steps will require elaborate analysis that might need some time of a post-doc or a Ph.D. candidate
3.8 Restructure breeding program to better capitalise on genomic selection

The immediate effect of genomic selection is the ability to rank young bulls and to use these rankings to pre-select young bulls for testing. Also genomic information can also contribute to the selection of bulls dams. However, follow recommendations on how the breeding program can be re-structured to fully capitalise on genomic selection can only be done once the genomic selection process is in place. Steps to implementation of breeding program restructure are:

a. Determine the accuracy of genomic prediction of young bulls and undertake selection of young persons for progeny testing
b. Determine how many young bulls to be tested given the accuracy of genomic predictions
c. Examine genomic predictions of females to aid selection of bull dams

3.9 Continual review and refinement of the breeding program.

All aspects of the breeding program should be continually reviewed and refined – as new data becomes available, as breeding objectives of the buffalo keepers change, or as technologies advance. This should incorporate periodic (e.g. 3 to 5 year) reviews amongst others:

a. The objective of the breeding program
b. The trait recording scheme for both nucleus and field progeny test animals, and the utility of the associated database
c. The analytical methodology for calculation of selection index / EBVs, including:
   i. Re-estimation of the various component estimates feeding into BLUP analysis, using the full complement of breeding program data
   ii. Re-estimation of the economic values of traits feeding into the selection index, using a profit model
d. Genomic prediction to be done routinely (about twice in a year) with reference population, and therefore prediction equations, updated as more data accumulates.
e. Overall performance of the breeding program, considering (amongst other) rates of genetic gain in individual traits and the selection index, as well as inbreeding levels. Programme shall be altered to achieve increased genetic gains (whilst maintaining breeding) or improved cost effectiveness.
f. Carrying forward the breeding program in relation to dissemination of the improved buffalo genetics
4. Concluding comments

The varied recommendations here will lead to use of the latest advances in animal breeding to be incorporated into the ICAR-CIRB led Murrah Buffalo breeding program. The sustainability of the above outlined approach is dependent on quantitative geneticists/animal breeders within the CIRB team having sufficient capacity to run the varied analysis, interpret the results in a critical manner, and correctly feed these results into breeding program implementation. To this end, ILRI will help build CIRB capacity through targeted workshops and back-stop support. Implementation of the recommendations will help to both increase the rate of genetic gain within the breeding nucleus, as well as reduce the cost of the breeding program.
## Annex 1. Additional details of the ICAR-CIRB led Murrah buffalo breeding program

<table>
<thead>
<tr>
<th>Lead centre</th>
<th>ICAR Central Institute for Research on Buffaloes (CIRB) Hisar, India</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other involved centres</td>
<td>ICAR institutes:</td>
</tr>
<tr>
<td></td>
<td>National Dairy Research Institute (NDRI), Karnal</td>
</tr>
<tr>
<td></td>
<td>Indian Veterinary Research Institute, (IVRI) Izatnagar</td>
</tr>
<tr>
<td></td>
<td>ICAR Research Complex for Eastern Region, Patna</td>
</tr>
<tr>
<td></td>
<td>Universities:</td>
</tr>
<tr>
<td></td>
<td>Lala Lajpat Rai University of Veterinary and Animal Sciences (LUVAS), Hisar</td>
</tr>
<tr>
<td></td>
<td>Guru Angad Dev Veterinary and Animal Sciences University (GADVASU) (Ludhiana)</td>
</tr>
<tr>
<td>FIELD UNITS (3):</td>
<td>CIRB, NDRI, GADVASU</td>
</tr>
<tr>
<td>Discontinued Centres:</td>
<td>Livestock Research Station (Mamnoor) [Discontinued from 2017]</td>
</tr>
<tr>
<td></td>
<td>KUVASU [Discontinued from 2014]</td>
</tr>
<tr>
<td></td>
<td>The main involved institutes are CIRB, NDRI, LUVAS, GADVASU and IVRI</td>
</tr>
<tr>
<td>Features of the Murrah breed</td>
<td>High milk production, in demand by farmers</td>
</tr>
<tr>
<td></td>
<td>Improver breed across the country and globe</td>
</tr>
<tr>
<td>Breeding objective</td>
<td>Increasing milk yield and reproductivity rates, whilst maintaining body confirmation/ type true to the Murrah breed</td>
</tr>
<tr>
<td>Type of breeding program</td>
<td>Open nucleus, dispersed across the six institutional herds</td>
</tr>
<tr>
<td></td>
<td>Progeny testing performed at 18-month intervals, involving both on-station and field testing.</td>
</tr>
<tr>
<td>Breeding herd</td>
<td>1,170 breedable females across the six institutional herds at March 2017</td>
</tr>
<tr>
<td>Recorded traits</td>
<td>Summary of recorded traits for institutional herds:</td>
</tr>
<tr>
<td></td>
<td>Milk: CIRB weekly milk yields, until drying or 305 days (a lactation record is not calculated if lactation length is less than 210 days). Milk measured each Friday and calves are not suckled on this day. Milk constituents tested once per month, done following birth date (e.g. animal born on fifth of the month, then testing every fifth of the following months). Other centres wean animals early and record daily yields (also without suckling).</td>
</tr>
<tr>
<td></td>
<td>Demographic events: births, deaths (including cause), sales etc.</td>
</tr>
<tr>
<td></td>
<td>Productivity - weights at various ages</td>
</tr>
<tr>
<td></td>
<td>Semen quality</td>
</tr>
<tr>
<td></td>
<td>Animals’ confirmation</td>
</tr>
<tr>
<td></td>
<td>Artificial insemination information</td>
</tr>
<tr>
<td></td>
<td>Health treatments</td>
</tr>
<tr>
<td></td>
<td>Disease testing (Tuberculosis, Johne’s disease, Brucella)</td>
</tr>
<tr>
<td></td>
<td>Data is collected into various well-organized registers; registers also used for recording of calculated/secondary data (e.g. lactation milk yield). A register also exists for feeding regime (at group level).</td>
</tr>
<tr>
<td></td>
<td>Summary of recorded traits for field progeny test animals</td>
</tr>
<tr>
<td></td>
<td>AI sire and dates; calving dates</td>
</tr>
<tr>
<td></td>
<td>Milk records on monthly basis</td>
</tr>
</tbody>
</table>
| Database | Institutional herds:  
Currently mainly in herd books (registers) with some secondary (derived) data in electronic format. The network improvement program is intending to have all data digitalized and in a common database (for all centres). There is discussion that this database be linked to herd management software.  
Field progeny test:  
Data entered into forms by field staff and then digitalized into word / excel tables by CIRB staff. |
| --- | --- |
| Selection of young bulls for progeny testing | Selected in ‘sets’ every 18 months  
12–16 bulls per set: each set has bulls from the different institutes.  
Bull selection by a committee of people from the various institutes, who also do a physical inspection of the bulls.  
Selection at around 2.5 years of age.  
Selection criteria:  
Mother giving more than 2,400 litres on first lactation; more than 3,000 litres on later lactations  
Bull is son of a progeny test bull’s progeny test bulls are used as the breeding program sires on elite dams.  
Murrah breed characteristics—shape of horn; tail has white section of no more than 17 cm; tongue is black  
Breeding soundness—libido, etc.  
Semen quality  
Mothers age at first calving, calving interval (less than 400 days)  
Skin thickness.  
Occasionally some bulls are introduced from outside e.g. government herds (such as Haryana government herd), commercial farms (such as Redhu), smallholder farms, if these are thought to be exceptional animals |
| Progeny test scheme – on station | Each bull selected for progeny testing is mated to a group of females, distributed across the farms, and the daughters (8–10 per young bull) from this mating are subsequently recorded for their first lactation milk yields between 3.75 and 4.75 years of age.  
To March 2017,  
156 bulls evaluated to eleventh set (over all centres);  
218 bulls will be tested by sixteenth set. |
| Progeny test scheme – on farm | Field testing is undertaken by three centres in adjoining villages: initiated in the year 2000. CIRB field testing—involves 10 villages, with approximately 6,000 animals inseminated per progeny test set with approximately 50% conception rate giving 3,000 animals born, 1,500 of which are female. Typically about 150 of these females (10%) will be recorded for milk, meaning between 9 and 13 records per bull. For bull set 11, there were a total (across all bulls) of 88 progeny. NDRI field testing—18–20 villages GADVASU field testing—25 villages Field staff are AI technicians who are contracted to the project—the field staff perform both AI and record keeping. Female animals are ear-tagged (previously were chipped, but there were difficulties in reading the chips). AI is performed at no cost to the farmers. Pregnancy testing by hand at three months (it requires lengthy permission process in India to have a mobile ultrasound) |
| Progeny testing – ranking of progeny tested bulls | For the tenth and eleventh bull set, Harvey model used to estimate breeding values for the progeny tested sires. Prior to this EBVs were not calculated—sires were ranked on their daughters lactation yield. Number of daughters with records per bull: 8–10 from institutional herds and 6–9 from field |
| Selection / mating in breeding herd | Since 2005, nominated mating is used with the two–three highest performing of the progeny tested bulls specifically mated to the 25–30% of the highest performing females, with the females ranked based on ‘305 days or less milk yield’ considered as the ‘standard lactation milk yield’ (SLMY). The remainder of the progeny test males are used for mating to the remainder of the females. Prior to 2005, there was no nominated mating—all of the progeny tested bulls were used for mating to the breeding females. Females are not actively selected, but culled if they have a problem such as disease, blind teats, infected teats, lameness, low milk production, or not able to become pregnant. |
| Dissemination from the nucleus | Semen for artificial insemination: On average, 10,000 semen doses are collected per bull (inclusive of both progeny tested and non-progeny tested), used for breeding program AI and commercial dissemination. Live animals: Males whose mothers produce, on average, more than 3,000 kg per lactation (or 2,500 kg for first lactation) and that are surplus to the breeding-program needs and sold to semen stations across the country. In 2016, 74 bulls were sold to semen stations. Males whose mothers produce, on average, more than 2000 kg per lactation are sold to the state government for natural mating, else to farmers. Indicative sale price: dependent on age and milk yield of dam. For example: 1–1.5 years old sold for INR 20,000–30,000 (higher price for better performing dams); breedable males sold for INR 50,000–70,000. |
| Other | The government is currently introducing a national identification scheme for buffalo and cattle. |
Incorporation of genomic selection into the Murrah buffalo breeding program