AgriPlex Genomics Bovine Parental Panel

Introduction

As of February 2022, the Food and Agricultural Organization states that livestock contributes 40% of all agricultural output and supports the livelihoods of almost 1.3 billion people globally (FAO, 2022). Since the Bovine (Bos Spp.) is the most populous non-human mammal, the cattle industry plays a critical role in feeding the world with the beef and dairy they produce. While cattle can be found all over the world, they mostly exist in India with 330 million cattle, and Brazil with 210 million cattle (Sahu et al., 2021). The two primary species of cattle that exist today are Bos Indicus (B. Indicus) and Bos Taurus (B. Taurus).

B. Indicus, also known as zebu, are native to the Indus Valley region (modern-day Pakistan and India.) These cattle are characterized by a fatty hump on their shoulders, a large dewlap, and drooping ears. Early domestication of zebu occurred in both India and Africa approximately 6000 YBP. In the early 20th century, B. Indicus were imported to Brazil, where they were considered "ecological" as they grazed freely on vast pastures, bred without husbandry, and produced lean meat without chemical residues (Brannstrom, 2004). They are appreciated for their adaptability to tropical climates and resistance to disease and parasites. B. Indicus are purposed as draught, riding animals, the sport of bull-taming, as well as for byproducts such as hides and dung for fuel and manure. While B. Indicus cattle produce beef and dairy, the B. Taurus subspecies produce the best quality food.

B. Taurus, also known as European cattle, make up most of the domestic beef and dairy cattle found throughout the world. They are thought to originate in Southwest Asia around 10,500 YBP (Bollongino, 2012). These cattle were primarily domesticated in Germany, Denmark, Hungary, Latvia, and the Netherlands. The European cattle was bred to generate large volumes of milk for human consumption and for quality tender meat with better flavor. B. Taurus, however, do not cope well with heat stress and tend to thrive in temperate weather. For this reason, cattlemen are interested in hybridizing B. Taurus cattle together with B. Indicus to combine environmental adaptability of the B. Indicus with the dairy yield of the B. Taurus. Cross species hybrids also display important heterosis effects (Madalena et al., 2017).

From the early stages of domestication, Bos species were subjected to selection and were bred into what is today's the beef industry, outputting meat for human consumption, the dairy cows industry for milk production, and more recently, the merged dairy-beef practices. The product quality for all bovine operations is dependent on their ability to improve and maintain the genetic make-up and genetic gains that accumulated over millenniums of breeding. A key tool to maintain bovine elite genetics is Parentage Testing.

Parentage Testing

Pedigree verification was initiated about 50 years ago with the analysis of blood groups (Stormont, 1967) and evolved into current parentage testing using Single Nucleotide Polymorphisms (SNPs) as DNA markers.

Parentage DNA testing allows to corroborate pedigree and the relationship between sires, dams, and offspring. SNP data in Parentage Testing is used in a process of elimination, where sires that cannot be the parents of a particular offspring are excluded, leaving the one (or few) remaining sires as the most likely parent for that offspring.

In turn, the relatedness and descent information obtained through the testing drives decision-making processes. For commercial producers, parentage testing drives herd management decisions. For example, when running multiple bulls in one pasture, knowing the most likely sire of each calf allows making the selection for future grouping of bull batteries based on calf performance. Additionally, producers using Artificial Insemination (AI) followed by natural service sires may have some calves with birth dates that are too ambiguous to determine the sire, and the only way to determine this correctly is by using parentage testing. This knowledge aids commercial producers who wish to retain only AI sired heifers as replacements or for embryo transfer calves.

Seedstock breeders rely on accurate pedigree information for Expected Progeny Differences (EPD) prediction. Considering the repercussions of the pedigree error rate, accuracy is crucial. A 10% pedigree error rate has been estimated to have a 6-13% effect on the inbreeding coefficient, 11-18% reduction on genetic trends in Estimated Breeding Values (EBV), and a 2-3% loss in the response to selection and causes a downward basis in heritability estimates. Missing sire data can have a large effect on the response to genomic selection and the variance of breeding value. While sire errors have a larger effect than missing sire information on genomic progress, their effect on genetic gain is additive (McClure et al., 2015). The initial standard SNP Parentage panel had 100 markers and was created by the International Society for Animal Genetics (ISAG). Another 100 SNPs were added to the panel (ISAG 200) as a supplement to be used in cases where finer resolution is required. The common practice is to conduct tiered testing, first with the ISAG 100 and to resort to genotyping the additional 100 SNPS when needed. Typically, more than three exclusions in the first testing tier led to excluding an animal as a potential parent of the calf in question. However, new research concluded that parentage should be determined with greater specificity and a larger number of SNP markers. The question of how many markers are needed to obtain a high probability that the parents, usually just the sire, are correct was addressed by the Irish Cattle Breeding Federation (ICBF) and reported in McClure et al (2015). 56,147 individuals representing more than 20 Bos taurus breeds including a mixture of Irish beef, dairy, purebred, crossbred, pedigree, and commercial male and females were genotyped with a core of 6909 SNPs from the Illumina LD chip.

The genotypic data that was obtained was used to calculate Minor Allele Frequencies (MAF), which served as the basis for how SNP markers were ranked. Panels with various numbers of SNPs (200, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1100, and 1250) were designed that contained the ISAG200 SNPs and increasing numbers of top MAF ranked SNPs. To test how many sires are

predicted at different SNP densities at the standard accepted 0.5-1% mis-concordance rate, 7,092 animals with known sire (of which 420 also had the dam SNP validated) were tested. The result from this analysis (Figure 1) indicated that 500 or more SNPs are needed to assure that only 1 sire will be predicted at <1.0% mis-concordance. These results concur with other reports (Strucken et. al. 2015, McClure, et. al. 2018 that using lower density parentage SNP panels, such as the ISAG 100 and ISAG200, can result in false-positive validations and result in multiple parents being predicted.

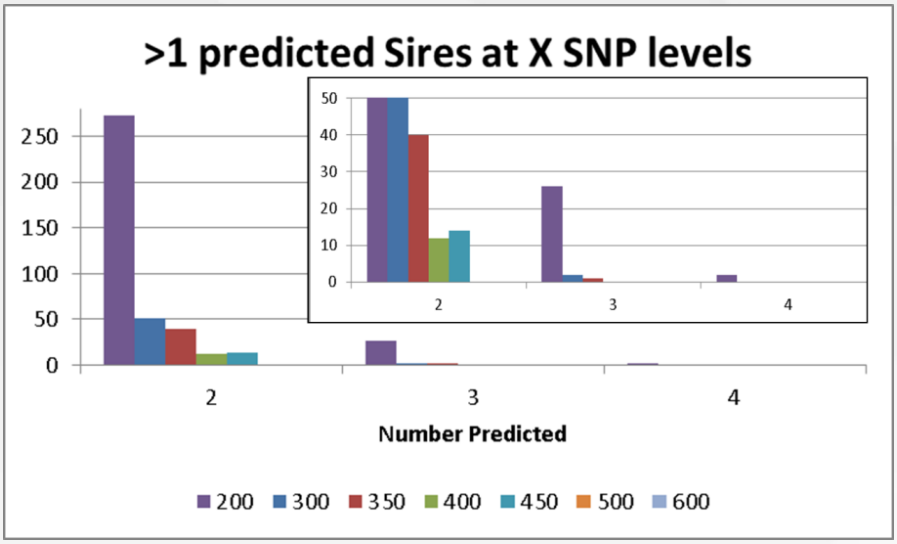

Figure 1 (Adapted from McClure et al, 2015): Count of predicted sires with ≤0.5-1% mis-concordances at varying SNP density levels.

The Parentage Panel

The Bovine Parental Panel is based with modifications on the ICAR SNP panel. The original panel consists of 554 SNPs and includes the 200 SNP recommended by ISAG for parentage verification in cattle, as well as an additional group of 354 SNP off these, 75 SNPS were added to aid in parentage discovery and are spread across chromosomes 1 to 29, while the remaining 279 SNP were selected from chromosomes 1, 2, 3, 5, 7, 8, 11, 13, 19 and 21. The SNPs were chosen out of the core markers that are in common to all Illumina bovine genomic chips available (Illumina 3K, LD, 50K, and HD SNP chips) and based on MAF and call rate.

With the Bovine Parentage Panel from AgriPlex Genomics, (Table 1), the SNP count was reduced by 27 markers. The reduction was made based on historical data pointing to markers whose MAF values were lower than initially estimated, or demonstrated performance problems that rendered some markers less informative (McClure et al., 2018).In addition, markers were evaluated based on call rate and performance in the PlexSeq multiplex – a key attribute at base of the economic efficiency of the panel.

The resulting 527 SNPs are distributed among 29 chromosomes with an average 6742 Mbp interval between markers and an average 19 SNPs per chromosome. The panel complies with recent literature, which states that more than 500 SNP are required to predict with confidence only 1 sire.

Chromosome	SNPs/ Chromosome	Average Inter SNP interval (Mbp)
1	51	3098
2	51	2673
3	40	2971
4	13	8681
5	42	2910
6	10	12202
7	37	2999
8	37	3111
9	9	11854
10	11	9470
11	37	2815
12	8	9948
13	36	2370
14	6	13982
15	10	6497
16	6	14178
17	11	7093
18	9	7818
19	35	1737
20	10	6439
21	33	2134
22	8	6816
23	6	8915
24	6	10912
25	6	7545
26	6	6346
27	7	5644
28	7	6391
29	6	7962
X	0	

The Genotyping Platform: *PlexSeq™*, the mid-density, multiplexed, SNP genotyping platform

Several attributes of the PlexSeq™ process contribute to its unique value as a genotyping platform:

- The proprietary multiplexing algorithm, *PlexForm™*: The software designs all possible primers around all requested SNPs. Artificial intelligence algorithms identify the optimal sets of primers that can be mixed in one PCR amplification reaction.
- Once the amplifications are completed, the amplicon mixture is equivalent to barcoded libraries produced from other NGS methods. The process is unique because the samples produce amplicon libraries that are equivalent in concentration and do not require any additional equalization steps. A mixture of all the libraries is subjected to one bead cleanup and are loaded onto the sequencer. The process saves time, plasticware, and expenses.
- The method requires only minute quantities of crude DNA that can be isolated from a variety of tissues, enabling a quick and inexpensive DNA isolation process to start the genotyping workflow.
- The *PlexSeq™* workflow consists of:
    - Crude DNA isolation
    - Primary PCR: highly multiplexed, low volume (3ul) PCR amplifications
    - Secondary, barcoding PCR amplifications
    - Pooling: barcoded amplicons are combined into one tube, purified and quantitated
    - Sequencing on an NGS sequencer
    - This relatively simple workflow is amenable to automation; all steps can be carried out on liquid handlers and high-capacity thermocyclers. This enables high-throughput genotyping

Once the sequencing is complete, a proprietary allele frequency-based genotype calling analysis software, *PlexCall™*, provides an automated sequencer to data workflow. This java-based software is tuned for each assay and is fully automated based on only the sequencing output files and a sample sheet indicating sample location on the plate.

Two other features make *PlexSeq™* a unique fit for molecular breeding and parentage testing. These applications typically require the genotyping of large number of individuals. AgriPlex Genomics' extensive collection of barcode combinations allows simultaneous sequencing of thousands of individuals, limited only by the sequencer's capacity. Similarly, molecular breeding may require the addition or substitution of only some of the SNP markers as the individual breeding program advances or among programs as the parent's genetic makeup or focus changes. Those changes in SNP composition are also required for QA applications as the diversity of germplasm changes. The fact that the panel is a collection of PCR primers not tethered to a surface (e.g.: chips) provides the flexibility to dynamically customize and alter the composition of the markers in the panel so it best fits the germplasm or the application. The Agriplex Genomics parental test NP panel is available as a service from AgriPlex Genomics and is also available as a kit to be used by in-house genotyping laboratories.

Applications: Parentage Verification and Parental Discovery

Based on the principals set forth by ICAR, parentage analysis, either parentage verification and/or parentage discovery should be conducted considering a set minimum proportion of the SNP panel; these markers should be genotyped and called for each animal involved in the parentage analysis process (i.e., animal and each potential parent). The most informative SNPs for parentage analysis are markers in which both the animal and parent in question are homozygous. A practical expectation is that based on the total number of SNP available for the animal and parent(s) in question, about one-third of available SNP for parentage analysis are informative. However, this proportion depends on the average minor allele frequency of the included SNP within the population of animals being considered. For a complete set of guidelines, please see: ICAR Guidelines for Parentage Verification and Parentage Discovery Based on SNP Genotypes (https://www.icar.org/Documents/GenoEx/ICAR%20Guidelines%20for%20Parentage%20Verification%20and%20Parentage%20Discovery%20based%20on%20SNP.pdf).

Conclusions:

AgriPlex Genomics have adapted the publicly available ICAR bovine parentage SNP panel. This effort combined pruning the panel of the least preforming SNPs and multiplexing the remaining markers to fit the PlexSeq Next Generation DNA Sequencing platform. The resulting mid- density, PlexSeq panel combines the economic benefits associated with the PlexSeq platform's high-level multiplexing capabilities and operational efficiencies, and the panel's size of 527 SNPs which satisfy the number of SNP markers needed to predict only 1 sire in most circumstances and herd sizes. The Bovine Parentage panel is available as a service at AgriPlex Genomics' lab or as a kit to run in one's own in-house laboratory.

References:

Food and Agriculture Organization of the United Nations (FAO.) Animal production. Retrieved April 29, 2022, from https://www.fao.org/animal-production/en

Sahu, B. K., Pargania, A., & Pati, A. K. (2021, January 20). *A population estimation study reveals a staggeringly high number of cattle on the streets of urban Raipur in India*. PLOS ONE. Retrieved April 28, 2022, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0234594

Brannstrom, C. (2004). Zebu's Elbows: Cattle Breeding the Environment in Central Brazil, 1890-1960. In Territories, commodities, and knowledges: Latin American Environmental History in the nineteenth and Twentieth Centuries (pp. 218–246). essay, Institute for the Study of the Americas.

Bollongino, R., Burger, J., Powell, A., Mashkour, M., Vigne, J.-D., & Thomas, M. G. (2012). Modern taurine cattle descended from small number of near-eastern founders. *Molecular Biology and Evolution*, *29*(9), 2101–2104. https://doi.org/10.1093/molbev/mss092

Madalena, F. E., Constable, P. D., Hinchcliff, K. W., Done, S. H., & Grünberg, W. Taurine Cattle. In *Veterinary Medicine (Eleventh Edition), 2017* (11th ed., pp. 657–715). Elsevier B.V.

McClure, M. C., Mccarthy, J., Flynn, J., Weld, R., Keane, M., O'connel, K., et al. (2015). "SNP selection for nationwide parentage verification and identification in beef and dairy cattle," in *Proceedings, International Committee For Animal Recording Technical Series, June 2015*, eds Z. Kowalski, N. Petreny, M. Burke, P. Bucek, L. Journaux, M. Coffey, C. Hunlun, and D. Radzio (Krakow; Rome: ICAR), 175–181.

International Society for Animal Genetics https://www.isag.us/animalgenetics.asp

Eva M. Strucken, Yan C.S.M. Laurenson and Gudrun A. Brockmann (2015) Go with the flow – biology and genetics of the lactation cycle. Front. Genet., 26 March 2015. Sec. Livestock Genomics, https://doi.org/10.3389/fgene.2015.00118

Snelling, W. M., Chiu, R., Schein, J. E., Hobbs, M., Abbey, C. A., Adelson, D. L., Aerts, J., Bennett, G. L., Bosdet, I. E., Boussaha, M., Brauning, R., Caetano, A. R., Costa, M. M., Crawford, A. M., Dalrymple, B. P., Eggen, A., Everts-van der Wind, A., Floriot, S., Gautier, M., Gill, C. A., … International Bovine BAC Mapping Consortium (2007). A physical map of the bovine genome. *Genome biology*, *8*(8), R165. https://doi.org/10.1186/gb-2007-8-8-r165

Matthew C. McClure, John McCarthy, Paul Flynn, Jennifer C. McClure, Emma Dair, D. K. O'Conneli and John F. Kearney (2018) SNP Data Quality Control in a National Beef and Dairy Cattle System and Highly Accurate SNP Based Parentage Verification and Identification. Front. Genet., 15 March 2018, Sec.Livestock Genomics https://doi.org/10.3389/fgene.2018.00084

International Society for Animal Genetics. (2012). Guidelines for cattle parentage verification based on SNP markers. http://www.isag.us/Docs/Guideline-forcattle-SNP-use-for-parentage-2012.pdf. Accessed 12/16/2016

ICAR Guidelines for Parentage Verification and Parentage Discovery Based on SNP Genotypes Prepared by: ICAR DNA Working Group (https://www.icar.org/Documents/GenoEx/ICAR%20Guidelines%20for%20Parentage%20Verification%20and%20Parentage%20Discovery%20based%20on%20SNP.pdf)