# Genomic selection in dairy cattle: Integration of DNA testing into breeding programs

**Jonathan M. Schefers*† and Kent A. Weigel***

*Department of Dairy Science, University of Wisconsin, Madison 53706; and

†Alta Genetics USA, Watertown, WI 53094

## Implications

Genomic selection offers many advantages with regard to improving the rate of genetic gain in dairy cattle breeding programs. The most important factors that contribute to faster genetic gain include

- A greater accuracy of predicted genetic merit for young animals.
- A shorter generation interval because of heavier use of young, genetically superior males and females.
- An increased intensity of selection because breeders can use genomic testing to screen a larger group of potentially elite animals.

By increasing the accuracy and intensity of selection and shortening the generation interval, the rate of genetic progress for economically important dairy traits can be approximately doubled.

**Key words:** dairy cattle, DNA testing, genomic selection

## Introduction

The advent of DNA sequencing and high-throughput genomic technologies has resulted in the discovery of a large number of single nucleotide polymorphisms (**SNP**) in cattle and other food animal species. Automated methods for SNP genotyping are now commercially available, and the use of dense SNP arrays that cover the bovine genome and that explain the majority of genetic variation in important traits has been proposed by an approach called genomic selection or whole-genome selection (Meuwissen et al., 2001). In practice, genomic selection refers to selection decisions based on genomic estimated breeding values (**GEBV**). These GEBV are calculated by estimating SNP effects from prediction equations, which are derived from a subset of animals in the population (i.e., a reference population) that have SNP genotypes and phenotypes for traits of interest. The accuracy of GEBV depends on the size of the reference population used to derive prediction equations, the heritability of the trait, and the extent of relationships between selection candidates and the reference population.

In dairy cattle breeding programs, genomic selection allows breeders to identify genetically superior animals at a much earlier age. In fact, animals that have been DNA tested can receive an accurate GEBV before they reach sexual maturity. Before the advent of genomic selection, artificial insemination (**AI**) companies relied on progeny testing to identify bulls with high genetic merit. Before progeny testing a young bull, the average estimated breeding value (**EBV**) of his sire and dam, which is commonly referred to as parent average, was used to select young bulls with the highest genetic merit and had an accuracy (reliability) of only 30 to 40%. In a progeny-testing scheme, a group of elite cows was identified as potential dams of young bulls (i.e., bull mothers). Progeny testing was necessary because most traits of economic importance in dairy cattle (e.g., milk production) are sex-limited and can be measured only in females. These bull mothers were mated to elite progeny-tested sires from the previous generation for the specific purpose of producing bull calves. Once these young bulls reached sexual maturity, which typically occurred at about 12 months of age, they were mated to a large number of cows on commercial farms, with the goal of producing approximately 100 daughters. Approximately 3 years later, the daughters of these young bulls would begin lactating, and this information was used to calculate the EBV of their sires for milk production and other key traits, which typically had reliabilities of 75 to 85%. At this point, these bulls were approximately 4.5 years of age, and the AI companies would decide which bulls should be culled and which bulls should be marketed to dairy farmers for the purpose of siring the next generation of replacement heifers. Overall, progeny-testing schemes are time consuming and costly because the AI companies have to wait many years to obtain genetic predictions with sufficient accuracy for making selection decisions, and in the meantime, hundreds of bulls are housed "in waiting" while phenotypes are measured on tens of thousands of their daughters. The objective of this review is to describe how genomics will affect genetic progress and breeding programs in the future.

## Factors Affecting the Rate of Genetic Progress

Four main factors affect the rate of genetic change in a population undergoing artificial selection. The classic equation for explaining the rate of genetic change, as described by Falconer (1989), is shown below:

$$\Delta G = \frac{ir\sigma_A}{L},$$

where $\Delta G$ is genetic change, $i$ is the selection intensity, $r$ is the accuracy of selection (or reliability of the EBV), $\sigma_A$ is the additive genetic standard deviation of the trait of interest, and $L$ is the generation interval. The rate of genetic progress can be described in detail for each of the 4 pathways of selection according to the sex of the parent and offspring.

### *Selection Pathway*

***Sires of Males.*** Sires of males (**SM**) represent the most elite males that are selected to be sires of the next generation of young bulls. This


doi:10.2527/af.2011-0032

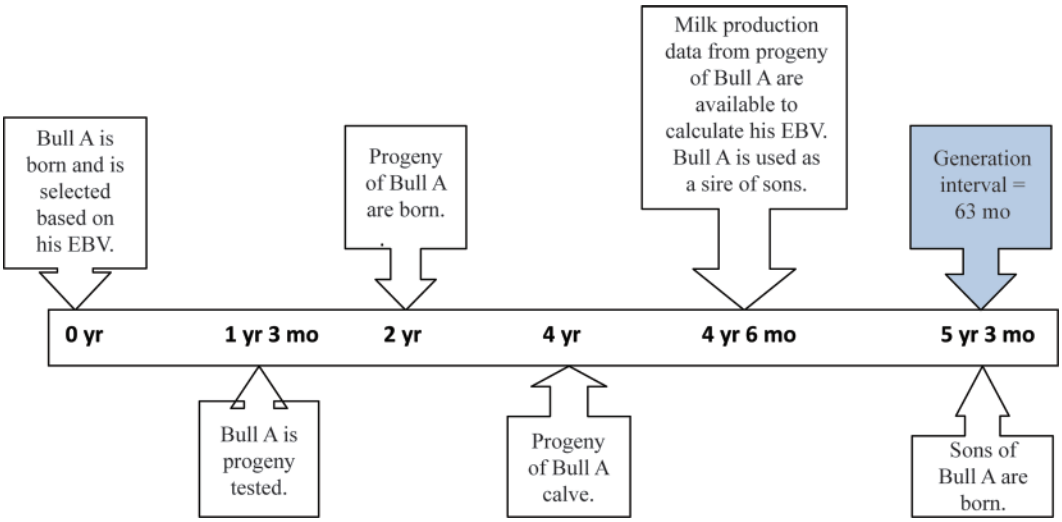


**Figure 1.** Timeline of a traditional artificial insemination breeding program based on progeny testing. EBV = estimated breeding value.

group is chosen based on EBV or GEBV and is typically composed of <5% of the males whose semen is marketed to dairy farmers. These bulls are often referred to as "sires of sons."

***Sires of Females.*** Sires of females (**SF**) represent a larger group of males that have been selected based on EBV or GEBV and whose semen is used to breed the general population and produce replacement females for commercial farms. These bulls are typically referred to as "active AI sires."

***Dams of Males.*** Dams of males (**DM**) represent a group of elite females that are selected based on EBV or GEBV and that usually rank among the top 1% of the population. These cows are mated to elite bulls from the SM group for the purpose of producing bull calves, and they are more commonly referred to as "bull mothers."

***Dams of Females.*** Dams of females (**DF**) represent the large population of females that are primarily used to produce milk rather than breeding stock. These cows, which are often referred to as "commercial cows," are routinely mated to bulls from the SF group to initiate lactation, resulting in the next generation of replacement heifers.

## Generation Interval

Because of the heavy reliance on progeny testing for sex-limited dairy traits, generation interval is the most important factor affecting the rate of genetic change in dairy cattle breeding programs. Generation interval ($L$) is defined as the average age of the parents when the progeny are born. Biologically, the shortest possible generation interval is the sum of age at sexual maturity and gestation length. This limitation can be circumvented by using advanced reproductive technologies, such as in vitro fertilization (**IVF**) of prepubertal heifers, but this practice is not commonly used in dairy cattle breeding programs.

As noted earlier, traditional progeny testing is a time-consuming process, and because breeders want highly reliable EBV when making selection decisions, the generation interval for the SM pathway is extremely long. Figure 1 shows the timeline for a traditional progeny-testing scheme, which has a generation interval for the SM pathway of approximately 63 months.

Genomic selection allows AI companies to make decisions based on GEBV, which are available at a very young age. Therefore, younger bulls can be used as sires of sons in the SM pathway, and the age at which they can be used is limited only by their sexual maturity. Instead of waiting a minimum of 4.5 years to use progeny-tested bulls as sires of sons, AI companies can use the best DNA-tested young bulls as sires of sons by roughly 1 year of age. This drastically reduces the generation interval in the SM pathway and, as noted by Schaeffer (2006), it could lead to doubling of the rate of genetic gain. Yearling bulls that have GEBV information but lack phenotypic data on their daughters are often referred to as "genomic bulls." There has been an immense shift among the AI companies toward the use of genomic bulls in the past 3 years. Some AI companies use almost all genomic bulls as sires of sons, whereas other companies use a combination of genomic bulls and progeny-tested bulls. Genomic bulls that are considered as sires of sons by AI companies have higher genetic merit, on average, but lower average reliability, so AI companies try to minimize risk by considering a larger number of genomic bulls as sires of sons. Figure 2 illustrates the timeline for an aggressive AI breeding program based on using genomic bulls as sires of sons. The key feature of this timeline is that if AI companies are very aggressive in using 1-year-old genomic bulls as sires of sons, the generation interval for the SM pathway can be reduced to 21 months. Estimation of the SNP effects for computing genomic predictions relies on the genotypes and phenotypes of reference animals, and in this case, young selection candidates are 3 generations removed from the most recent data on progeny-tested ancestors in the reference population. In the example shown in Figure 2, bull B is 54 months of age when phenotypes of his daughters become available for computation of his EBV, and by this time, breeders are already using semen from his best grandsons to produce his great-grandsons and great-granddaughters.

Breeders are also using GEBV information to select females at a much earlier age. Before the introduction of genomic selection, breeders generally waited until cows were at least 2 years of age before using reproductive technologies, such as multiple ovulation and embryo transfer

(**MOET**) or IVF. However, DNA testing of heifer calves provides GEBV with sufficient accuracy to allow selection decisions to be made before heifers reach sexual maturity, and the most aggressive breeders are making heavy use of virgin (yearling) heifers as bull mothers. This strategy can reduce the generation interval in the DF pathway to 21 months as well.

### Selection Intensity

Two factors affect selection intensity ($i$) in a breeding program. First, selection intensity is dependent on the size of the population. Greater selection intensity can be achieved in large populations because more selection candidates can be screened in search of genetic "outliers." Second, selection intensity is dependent on the proportion of animals selected from the population to serve as parents of the next generation. In the era of genomic selection, DNA samples can be taken shortly after birth, and AI companies can screen thousands of bull calves to find a few elite individuals for their breeding programs. In practice, most North American AI companies will DNA test and subsequently discard 10 to 20 genetically inferior bull calves for every elite young bull that enters the program. Since the start of genomic testing in 2009, approximately 33,000 young bulls have been DNA tested in North America. Since January 2011, approximately 1,000 bulls per month have been tested in North America (G. R. Wiggans, US Department of Agriculture, Agricultural Research Service, personal communication). Potential bull dams can be tested as well, and this gives AI companies and breeders an opportunity to screen thousands of potentially elite cows and heifers on commercial farms for the purpose of identifying a few superior individuals that can be propagated by MOET or IVF.

### Accuracy of Selection

In a progeny-testing program, the accuracy of selection ($r$) depends largely on the number of offspring per sire and, hence, on the number of cows in progeny test herds that are available for mating to young, unproven bulls. With genomic selection, accuracy is primarily a function of the size of the reference population that is used to estimate SNP effects, which in turn are used to compute GEBV of selection candidates. This reference population may consist of genotyped females that also have phenotypes, genotyped males that have daughters with phenotypes, or a combination of the two. At present, the reliabilities of GEBV for production traits are often 70% or greater in North American Holsteins (Van Raden et al., 2009), which is twice the level of reliability associated with traditional parent averages that are computed from pedigrees.

### Genetic Variation

The genetic standard deviation ($\sigma_A$) reflects the underlying genetic variability of a given trait within the population. Inbreeding decreases the effective population size, which can reduce the amount of genetic variation available for selection and reduce the rate of genetic progress. However, in comparison with other factors in the equation for genetic progress, relatively little can be done to increase the amount of genetic variation within a population.

## Application of Genomic Selection in Females

### Genomic Testing

Many progressive breeders are using genomic testing for the majority of their cows and heifers to identify those females that received the most favorable combination of genes from their parents. Currently, a low-density (**LD**) chip with 6,909 SNP (Illumina, 2011a) and a medium-density (**50K**) chip with 54,609 SNP (Illumina, 2011b) are the products used most frequently by breeders, and GEBV for production, health, and conformation traits can be computed using genotypes from either chip. Recently, the LD chip replaced the 3K chip with 2,900 SNP (Illumina, 2001c) because of greater gains in reliability and improved readability among SNP genotypes. The main difference between the LD chip and 50K chip is cost, and the LD chip is more affordable for breeders who wish to genotype a large number of cows, heifers, or calves. To combine information from SNP chips of different densities when calculating genetic evaluations, im-

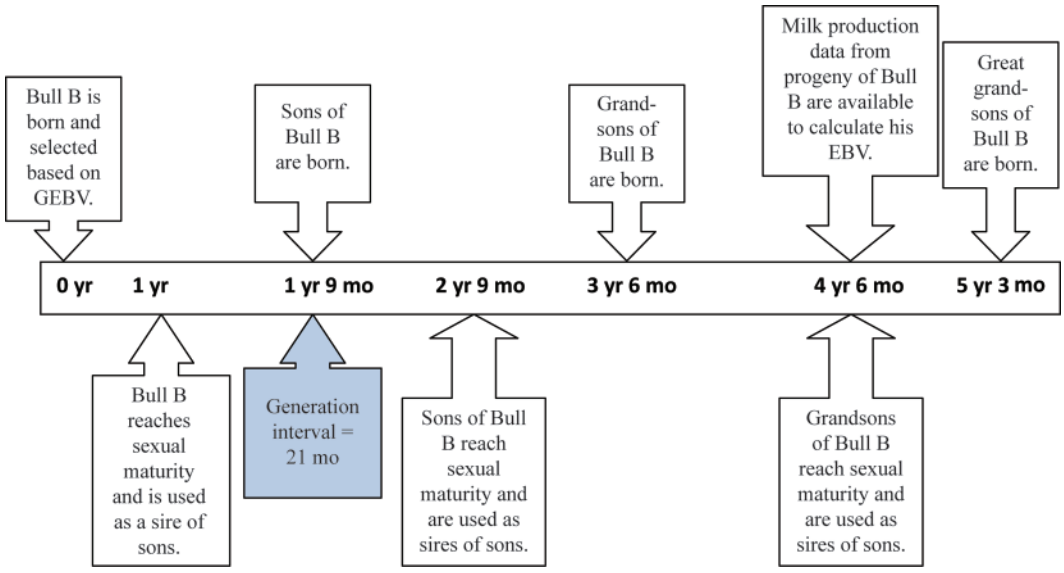


**Figure 2.** Timeline of an aggressive artificial insemination breeding program based on the use of genomic bulls as sires of sons. GEBV = genomic estimated breeding value; EBV = estimated breeding value.

putation is used to extend the LD subset of markers to the full 50K set. Imputation refers the prediction of genotypes for missing markers based on family information or linkage disequilibrium. A disadvantage of the LD chip is that imputation errors can occur, and this can lead to a slight reduction in the accuracy of GEBV. Dassonneville et al. (2011) reported error rates ranging from 2 to 6% when imputing 50K genotypes from LD genotypes.

Genomic testing services are currently offered by breed associations, AI stud services, and some privately owned companies. The LD chip costs $43 to $55 per animal, and reliabilities of the resulting GEBV for production traits in Holsteins are approximately 60 to 65%. The 50K chip is more costly, at $125 to $135 per animal (Holstein Association USA, 2011), but reliabilities for production traits in Holsteins are roughly 70%. For elite females that are likely to be bull dams or embryo donors, breeders often prefer the 50K chip because of its greater reliability.

### *Advanced Reproductive Technologies*

Once a breeder identifies genetically superior females using genomic testing, and when these animals reach sexual maturity, they usually become part of a MOET-based program. Factors such as time, expense, and number of available recipients affect the extent to which a breeder decides to invest in MOET. The most aggressive breeders use a combination of MOET and IVF for their top animals. Typically, elite females undergo their first embryo transfer at 12 months of age. If these females produce a sufficient number of viable embryos, they can be superovulated 3 times before they reach 15 months of age, at which time they are inseminated using a conventional AI service. Pregnancy can be diagnosed at 30 days of gestation, and after they are diagnosed as pregnant, these heifers can enter an IVF program. The IVF collection, typically called an "aspiration," is the process of harvesting unfertilized oocytes directly from the ovaries of the donor animal. These oocytes are fertilized in vitro 1 day after collection. Once fertilized, the embryos are cultured and grown for 7 days of incubation before transfer into recipient females. The main advantage of IVF is that once an animal has been confirmed pregnant, oocytes can be aspirated every 2 weeks. Donors can be aspirated safely between 30 and 100 days of gestation, so if a pregnant female is aspirated at 30 days of gestation, it is possible to carry out 6 aspirations before she reaches 100 days of gestation. Therefore, not only is it possible to create a large number of pregnancies from 1 donor, but it is also possible to mate this donor to 10 different sires before her first calving (3 superovulations prebreeding, 1 conventional AI breeding, and 6 IVF collections during her pregnancy). In extreme cases, some breeders have put highly superior females into a continuous MOET and IVF program. For breeders who want to generate as many pregnancies as possible from a single donor, IVF is the most efficient approach. It is possible that a donor animal might never have a natural calf. However, if a breeder wants to continue to make genetic progress, a continuous MOET and IVF program for an individual donor is impractical, because over time we would expect this donor to be displaced by a younger, genetically superior female.

### *Sire Selection*

The use of young genomic bulls by AI companies as sires of sons (SM pathway) or by breeders as sires of replacement females (SF pathway) continues to increase in popularity. Among dairy producers, there has been a major shift toward the use of genomic bulls in the SF pathway. Between 2006 and 2010, the total number of units of semen sold from young dairy sires increased by 13% (Olson et al., 2011), and the increased acceptance of genomics among dairy cattle producers has allowed extensive marketing of genomic bulls. Moreover, König et al. (2009) reported that breeding programs that use young genomic bulls would have greater profit than those that rely on conventional progeny testing, provided that at least 20% of the inseminations were to genomic bulls that lacked daughter records.

On many farms, it is becoming the norm to breed virgin heifers to genomic bulls because, on average, these yearling heifers have greater genetic merit than the lactating cows. As noted earlier, using young, genome-tested males in the SM and SF pathways and using young, genome-tested females in the DM and DF pathways can reduce generation intervals in these pathways by 50% or more. Because the reliability of GEBV from DNA testing is usually less (approximately 70%) than the reliability of EBV from progeny testing (about 85%), breeders tend to use a larger group or "team" of young bulls to mitigate risk. The advantage of a team-based approach to sire selection is that reliability of the average GEBV for a team of young bulls is considerably greater than the individual reliability of each bull. The formula for calculating the reliability of a team of bulls is given below:

$$\text{Team reliability} = 1 - \frac{1 - \text{average reliability of the individual bulls in the team}}{\text{number of bulls in the team}}.$$

For example, if a breeder uses a team composed of 5 young bulls with reliabilities of 70% for individual GEBV, the average GEBV for this team would have 94% reliability.

## Application of Genomic Selection in Males

### *Genomic Testing*

Since the introduction of genomic selection, the percentage of young dairy bulls that have been DNA tested is greater than 90% for the Holstein, Jersey, and Brown Swiss breeds (Olson et al., 2011). All young bulls that are considered for purchase by the major AI stud services are selected based on the results of genomic testing. Therefore, genetic gain is maximized by screening a vast number of bull calves because this increases selection intensity. The limiting factors with respect to testing more bull calves are time of the breeder and sire analyst, cost of the DNA test, and the willingness of the breeder to feed and house a large number of young bulls while waiting for their initial GEBV results.

Because genomic selection gives AI companies the ability to carry out accurate selection decisions at a young age by using DNA testing, these companies may decide to cut costs by purchasing fewer bulls, knowing that only males with the highest genetic merit will be marketable. In fact, there have been discussions among scientists and practitioners about eliminating progeny testing entirely. This could reduce sire development costs by up to 92% (Schaeffer, 2006) because the biggest costs associated with progeny testing are housing and feeding. However, it is unlikely that AI companies will eliminate progeny testing in the short term for 2 reasons. First, a global market for progeny-tested dairy bulls still exists, largely because genetic evaluations for these bulls have higher reliability. Second, ongoing measurement of daughter phenotypes is essential because prediction equations for calculating GEBV (i.e., SNP effects) should be updated periodically to increase the size of the reference population, ac-

count for changes over time in herd management practices or the genetic background of mates, and maintain high levels of linkage disequilibrium between the reference population and the selection candidates. Misconceptions exist that genomics eliminates the need for milk recording and type classification services. However, it is essential to keep the reference population current through continuous and consistent data recording so that genomic predictions are not based on a reference population that is many generations removed from the current milking dairy cattle population.

### Selection of Bull Mothers

Before genomic selection, breeders and AI companies tended to select potential bull mothers based on a multiple-trait index of production, health, and conformation traits. However, to be considered as bull mothers, these cows would need complete pedigrees and would have to meet minimum standards for phenotypic production and physical conformation. Cows that failed to meet these minimum criteria were excluded as bull mothers, and this significantly reduced the number of eligible cows. The movement to genomic selection has dramatically increased the pool of females that can be considered as potential bull mothers. It is no longer necessary to discard every female that has less-than-perfect pedigree, production, or conformation data. Furthermore, cows on large commercial dairy farms that focused primarily on the sale of milk were rarely considered as bull mothers because their pedigrees lacked several consecutive generations of high-ranking sires, or because herd management conditions favored strong performance of the entire group (i.e., pen or string) rather than extreme performance of selected animals within the group. In fact, management conditions in these large commercial herds might more closely reflect the conditions under which future daughters of these bull calves will be expected to perform, as compared with herds from which bull calves were typically purchased historically.

Another important consideration is that breeders and sire analysts can capitalize on the tremendous variation that occurs in each mating because of Mendelian sampling of alleles from the sire and dam. For example, a particular cow may not have an EBV or GEBV that is sufficient to reach elite status, but when mated to an elite sire, her calf might inherit a highly favorable combination of alleles that will warrant his purchase by an AI company or her use as an IVF donor. Calves of this type that are recognized as superior through DNA testing can contribute more than just high genetic merit; they can also enhance genetic diversity in future generations by virtue of their unique pedigrees and genetic backgrounds.

## Areas of Concern

### Inbreeding

A key concern among dairy breeders is the difficulty in maintaining genetic diversity when mating elite females and males because of genetic relationships among these individuals. It has been suggested that the rate of inbreeding might decrease in the era of genomic selection because a greater number of bulls are being used as sires of sons. A good description of the potential impact of genomic selection on the rate of inbreeding was given by Hayes et al. (2009):

> Consider the selection of young bull calves to become part of a progeny test team. In the absence of genomic information, and because the young calves do not have any daughters, their breeding value is predicted as the average of the breeding value of their sire and dam. Two full sibs therefore receive the same breeding value, and if this is high enough, they will both be selected to form part of the progeny test team. If genomic information is available, the Mendelian sampling term (the result of the sampling of the sire and dam alleles during gamete formation) is captured and 2 full sibs receive different breeding values, and may not both be selected to form part of the team, which leads to a decrease in the rate of inbreeding. However, if the generation interval of the breeding program is halved to take advantage of the accurate GEBV available at birth, the resulting increase in inbreeding per year may be greater than the decrease from capturing the Mendelian sampling term. (pp. 439–440)

Because genomic selection will allow the generation interval to be minimized in dairy cattle breeding programs, it is likely that the rate of inbreeding per year will increase. Furthermore, although the major AI companies consider a large number of genomic bulls as sires of sons, most of these young bulls have many ancestors in common, including sires and maternal or paternal grandsires. In addition, most of the elite young males are closely related to the elite young females in the same population because both were created from the same donor dams in MOET and IVF programs. For this reason, it is difficult to minimize inbreeding when mating elite animals. In the future, tools for routinely monitoring the genetic diversity of existing animals and managing the expected diversity of their future progeny by using genome-based mating programs may help to address this challenge.

### Bias

Patry and Ducrocq (2011) investigated the role of genomic preselection in young bulls. Because young bulls are selected based on high GEBV, they typically have superior Mendelian sampling contributions. Conversely, young bulls that are culled because of low GEBV typically have inferior Mendelian sampling contributions. Because these selection and culling decisions occur before the young bulls have an opportunity to sire any progeny, the assumption of random Mendelian sampling in genetic evaluations is violated. In the future, this could lead to bias in the GEBV of young bulls and heifers; therefore methods that account for bias attributable to genomic preselection should be investigated and incorporated into genetic evaluation programs.

## Conclusions

Genomic selection already plays an important role in dairy cattle breeding programs, and this will be the case for the foreseeable future. Increases in the accuracy of genetic predictions for young animals will dramatically decrease the generation interval, and when coupled with opportunities to increase selection intensity, the rate of genetic progress in dairy cattle will increase significantly. Many breeders have embraced genomic selection and routinely use GEBV when purchasing semen or deciding which cows and heifers merit investment in reproductive technologies such as MOET and IVF. At the same time, AI companies are aggressively using genomic testing when determining which young bulls to purchase, marketing semen to dairy producers, and identifying elite females that can make positive genetic contributions to the next generation.

DOCKET ALARM

# Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

## Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

## Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

## Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

## API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

**LAW FIRMS**
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

**FINANCIAL INSTITUTIONS**
Litigation and bankruptcy checks for companies and debtors.

**E-DISCOVERY AND LEGAL VENDORS**
Sync your system to PACER to automate legal marketing.