

Quantifying breed composition through genomic information

By Jason Reding, Marike Hillocks, Hannah Kruger and Dr Bobbie van der Westhuizen

For centuries, breeders only relied on pedigree records and visual appearance to determine an individual animal's breed composition. While these remain valuable tools, they do not always tell the full story and can be subject to different interpretations.

Pedigree records may be incomplete or contain errors, and although animals can look similar, they often have very different genetic backgrounds. Advances in genomic technology now make it possible to examine an animal's genetic make-up, enabling the estimation of its breed composition with far greater confidence.

Every animal inherits a unique DNA pattern from its parents. However, generations of selective breeding have resulted in animals of the same breeds sharing many common DNA characteristics. By comparing an animal's DNA with that of thousands of animals from known breeds, it is possible to estimate the proportion of each breed present in its genetic make-up.

A reliable genomic reference

Breed prediction relies on reference populations, which are large groups of animals whose breed composition has already been confirmed. Their genomic information acts as a benchmark against which other animals are compared. The larger and more diverse the reference population, the more accurately breed composition can be estimated, as it captures the natural variation within each breed while preserving the characteristics that make each breed unique.

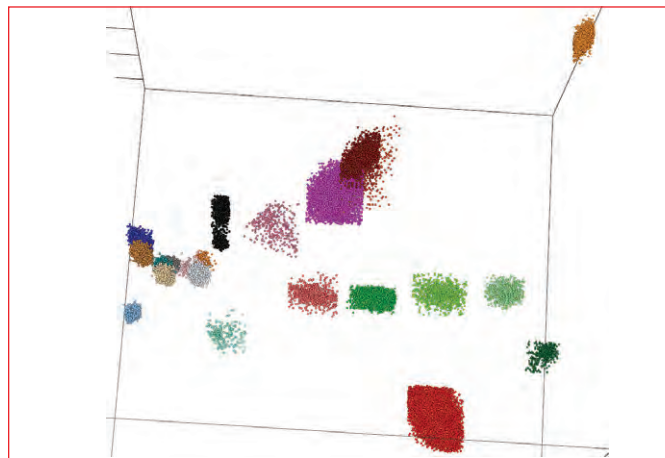
For this reason, the reference population is carefully assessed to ensure that each breed is represented by animals that reflect its genetic characteristics. Once the reference population has been established, the next step is to compare the test animal's DNA with this benchmark.

Modern genomic testing evaluates thousands of genetic markers spread across the entire genome rather than only a few locations in the DNA code. These markers, known as single nucleotide polymorphisms (SNPs), each provide a small piece of information about the animal's ancestry. Statistical methods are then used to combine information from all markers to estimate an animal's breed composition, using SNP-BLUP to obtain the best possible estimate.

Evaluation of reference populations can also be done using another statistical method called principal component analysis and visualising the results within a three-dimensional 'genebox'. The combination of an accurately established reference population, thousands of genomic SNP markers, and advanced statistical analysis provides a powerful tool for estimating breed composition for each genotyped animal.

Instead of relying solely on pedigree records or physical appearance, genomic analysis examines the DNA itself to obtain an

Figure 1: The 3D 'genebox', based on SA Stud Book's Breed Key version 2.



The box shows the genetic distances between animals (represented by dots) and how each animal clusters within its breed (each colour represents a different breed).

objective estimate of an animal's genetic background, even when pedigree information is incomplete or uncertain. Even in cases of mixed-breed origins, this method is extremely accurate at tracing the roots of genotyped animals by quantifying the percentage of each breed that contributed to their genetic composition.

Expanding genomic capability

The current Beef Genomics Program 2, or BGP2, has been a resounding success in building reference populations for numerous breeds using SA Stud Book's genomic services. Breed Key version 1 utilised three dairy breed and eight beef breed reference populations, using 24 000 SNP markers to estimate the genetic breed composition of a genomically tested animal. Breed Key version 2 has been upgraded to include 11 more beef breeds in conjunction with utilising 41 000 SNP markers, allowing for a more comprehensive assessment of an animal's true genetic make-up. The rollout of Breed Key version 2 occurred in June 2026, with breeders already receiving the breed composition of their most recently genomically tested animals.

It is important to remember that breed composition estimates are exactly that: estimates. They represent the animal's genetic similarity to the breeds represented in the reference population. As more animals are genotyped and added to the reference population, and as the quality of that reference population improves, these estimates become more accurate and reliable. **SF**