agriculture

MDPI

Article

Kinship and Network Analysis of Two South African Beef Cattle Breeds Using Pedigree and High-Density SNP Markers

Khulekani S. Khanyile [1,2,*], Azwihangwisi Maiwashe [1], Nozipho A. Magagula [3], Este van Marle-Köster [2] and Avhashoni A. Zwane [4,*]

[1] Animal Breeding and Genetics, Agricultural Research Council—Animal Production, Irene 0062, South Africa; norman@arc.agric.za
[2] Department of Animal Science, University of Pretoria, Hatfield, Pretoria 0028, South Africa; este.vanmarle-koster@up.ac.za
[3] Department of Animal Sciences, Tshwane University of Technology, Pretoria 0001, South Africa
[4] Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Hatfield, Pretoria 0002, South Africa
* Correspondence: kskhanyile@gmail.com or khanyilek@arc.agric.za (K.S.K.); avhashoni.zwane@up.ac.za (A.A.Z.)

Abstract

Accurate genealogical records are essential in livestock breeding for maintaining genetic diversity, preventing inbreeding, and mapping of economically important traits in beef production. This study aimed to assess parent–offspring relationships within South African Bonsmara and Nguni cattle populations using both traditional pedigree records and genomic data. Hair samples from 119 Nguni and 311 Bonsmara cattle were genotyped using the BovineSNP50 array, and these were imputed to Illumina BovineHD BeadChip using updated SNP coordinates from the assembly genome (ARC—UCSD 1.2). Quality control and data filtering were performed using PLINK v1.9, while relationship inference was conducted using KING v2.2.8 and PLINK v1.9 software for principal component analysis, IBD metrics and Mendelian error-based exclusion. Categories of relatedness through network relationship analysis revealed a predominance of half-sibling relationships in both breeds, with 2317 such relationships identified in Nguni and 1221 in Bonsmara. Inference of parent–offspring pairs showed discrepancies with the recorded pedigrees, with 49 inferred pairs compared to 47 recorded pairs in Nguni, and 62 inferred pairs compared to 75 pairs recorded in Bonsmara. Relationships based on IBD using PLINK with a 'PI-HAT' threshold greater than 0.45 revealed unique parent–offspring inferences that differed from those obtained using KING v2.2.8. Phylogenetic network analysis assigned each individual's genomic origin independent of the pedigree records, supporting the efficiency of SNP data for genetic assignment. These results demonstrated that SNP-based pedigree verification can accurately identify parent–offspring and half-sibling relationships, providing a reliable foundation for recombination analysis and supporting precise trait mapping and informed selection in breeding programs.

Keywords: pedigree; single-nucleotide polymorphism; Mendelian error; identity-by-descent; breeding

Academic Editors: József Rátky, Manuel García-Herreros and Pedro Manuel Aponte

Received: 22 October 2025
Revised: 13 February 2026
Accepted: 16 March 2026
Published: 19 March 2026

1. Introduction

The field of animal genetics has undergone a significant transformation due to the advent of high-throughput sequencing and genotyping technologies, particularly the use

of single-nucleotide polymorphism (SNP) arrays. Single-nucleotide polymorphisms are abundant, stable, and uniformly distributed across the genome, making them ideal markers for both population-level analyses and precise kinship estimation [1]. The practical use of genomic tools for kinship assessment in local cattle populations remains limited, especially in the smallholder sector. Most studies on indigenous beef cattle in South Africa have focused on population structure, breed characterization, or admixture analyses [2–6], with few attempting to reconstruct or verify within-breed kinship in the absence of reliable pedigree data. Moreover, widely used computational algorithms for relationship inference often assume random or genetically homogeneous populations, which can lead to biased results when applied to structured or admixed populations such as those found in communal farming systems [7].

Livestock production is indispensable in the global agricultural economy and remains a key contributor to household livelihoods, food security, and socioeconomic development, particularly in low- and middle-income countries. In South Africa, cattle are highly valued not only for their sociocultural significance but also because they are a primary source of animal protein and income for farmers [8,9]. Indigenous breed such as the Nguni and Bonsmara, a composite breed, are widely distributed across diverse agro-ecological zones and have been strategically bred for traits such as disease tolerance, adaptability, and productivity [10,11]. These breeds have been reported to have originated from crossing taurine and indicine to form a Sanga (*Bos Taurus africanus*) breed from the East-Africa and found a niche in Southern Africa through human migration [6,12]. The successful implementation of breeding and genetic improvement strategies depends fundamentally on the availability of accurate genealogical or pedigree data. Accurate kinship estimation is critical not only for controlling inbreeding but also for improving selection decisions and ensuring long-term genetic gain.

The probability that two alleles in an individual are identical by descent (IBD) can be used to investigate the level of inbreeding in a population [13]. In addition, genomic relationship matrices, which quantify the overall genomic similarity between pairs of individuals, can be used to accurately reconstruct pedigrees with or without relationship records. Single-nucleotide polymorphism data can be used to identify Mendelian errors among recorded parent–offspring pairs, providing an effective quality control measure when only partial lineages are available in a population [14]. SNP data provides a powerful, marker-based alternative to infer genetic relationships, population structure, inbreeding, QTL mapping and demographic history, offering insights that would otherwise be impossible to obtain. For construction of pedigrees through SNPs markers, it is possible to establish relationship clusters through phylogeny and network analysis, allowing for a more precise assignment of individuals to their lineage of origin [15] and to classify pairwise relationships into well-defined kinship categories.

The ability to validate pedigree information and infer kinship using SNP data, even in the absence of genealogical records, holds great potential for beef cattle breeds kept by smallholder farmers where pedigree and performance recording are often limited or non-existent. By demonstrating the extent of pedigree misclassification and highlighting the utility of genomic information in resolving complex relationships, this study provides valuable insights for managing inbreeding and designing more efficient and genetically sustainable smallholder breeding programs where animal recording is limited. The aim of this study was to use both pedigree records and high-density SNP data to reconstruct pedigrees and verify kinship relationships for two Sanga breeds, Nguni and Bonsmara beef cattle of South Africa.

2. Materials and Methods

2.1. Animal Selection and Sample Collection

This study used the available pedigree data from the Agricultural Research Council (ARC), Integrated Registration and Genetic Information System (INTERGIS), and the DNA profiling databases to select Bonsmara and Nguni animals for genotyping. Hair samples were obtained from the ARC-Animal Production (ARC-AP) Biobank. This study was conducted following approval by the ARC-AP Ethics Committee (APAEC 2020/17) and University of Pretoria Ethics Committee (NAS374/2022).

A total of 311 Bonsmara animals were selected from different families originating from two temperate regions (Gauteng and Northwest provinces). Based on the pedigree information, 52 half-sib families with identifiable sire–offspring pairs were included in the selection. Bonsmara animals consisted of multi-generational relationships including grandparents, parents and offspring relationships. For Nguni population, a total of 119 parent–offspring pairs were selected from three farming districts in the temperate coastal region (Eastern Cape). The average half-sib family sizes for Nguni and Bonsmara were 2.8 and 6.0, respectively.

2.2. DNA Extraction and Genotyping

Hair samples from 311 Bonsmara and 119 Nguni animals were sent to the Agricultural Research Council—Biotechnology Platform (ARC—BTP, Pretoria, South Africa) for genomic DNA extraction and genotyping using the BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) [16]. DNA samples were quantified using Qubit v3.0 spectrophotometer to meet DNA concentration of $\geq$50.0 ng/µL for high-quality and appropriate quantities of DNA before SNP genotyping. Samples were genotyped in two groups per breed, one group on BovineSNP50 BeadChip (90% of the total samples, i.e., 103 Nguni and 279 Bonsmara) and the second group on Illumina® BovineHD 777K BeadChip (10% of the total samples, i.e., 16 Nguni and 32 Bonsmara), to allow for imputation. Illumina Genome Studio v2.0.5 software was used for SNP calling, and data was converted into PLINK v1.9 input format [1].

2.3. Data Imputation

Imputation was performed using Beagle v5.4 with genotypes from Nguni and Bonsmara cattle. A subset (16 Nguni and 32 Bonsmara) of animals including grandparents, parents, and offspring in the Bonsmara population, and parent–offspring pairs in the Nguni population, was selected as reference individuals for imputation to high density, using samples genotyped with the Illumina BovineHD 777K SNP panel. About 10 percent of the population was used as reference individuals in both breeds.

Imputation was conducted from a lower-density SNP50K panel to a 777K panel using Beagle's [17] default parameters as follows:

The analysis was executed with a set of optimized parameters, which were specifically reduced to improve performance and minimize computing time. Haplotype estimation and genotype phasing were conducted over a burn-in period of 6 iterations and 12 main iterations, with the Hidden Markov Model (HMM) for phasing configured to use 280 model states (phase-states). For the subsequent imputation of ungenotyped markers, the HMM was set to utilize 1600 model states (imp-states). This imputation process required a minimum cM length of 6.0 cM (imp-segment) for haplotype segments to be incorporated into the HMM statistics. Short Identity-by-State (IBS) segments were detected using a cM step length (imp-step) of 0.1 over 7 steps (imp-n steps). Furthermore, to manage marker density, individual markers were aggregated into a single unit if the maximum cM distance between them did not exceed 0.005 (cluster). General settings included an estimated

effective population size (ne) of 1,000,000 for the phased data and a maximum cM length of 400 cM for the sliding window (window). Finally, the output files were configured to include both the probability of alleles 1 and 2 (ap = true) and the genotype probability (gp = true).

For the updated assembly coordinates, a web-based version of the University of California, Santa Cruz (UCSC) LiftOver program URL https://genome.ucsc.edu/cgi-bin/hgLiftOver (accessed on 23 November 2023) was used to update genome coordinates from *Bos_taurus*_UMD3.1 to the latest assembly coordinates of ARS-UCD1.2 using imputed data containing 732,570 SNP markers, from the initial 772,071 SNPs markers before imputation. This study applied a simplified approach for imputation accuracy such as the use of the Concordance Rate (CR), calculated by comparing the imputed marker genotypes against the actual genotypes present in the original 50K panel. The CR represents the proportion of correctly imputed genotypes at these masked sites. From the initial and final number of markers, the imputation accuracy was estimated at 94.9%. The remaining marker set was subjected to quality control for downstream analysis.

2.4. Data Filtering and Quality Control

Using PLINK v1.9, quality control analyses for genotype data were performed on 119 Nguni samples with 485,236 SNPs and 311 Bonsmara samples with 465,097 SNPs. Genome-wide Mendelian inconsistencies were tested on all samples, and individuals with more than 5% uncalled genotypes (call rate < 0.95), low minor allele frequencies (MAF < 0.05), and a *p*-value less than 0.0001 for the Hardy–Weinberg equilibrium test were removed for further analysis. SNPs that had duplicate genomic positions were also excluded.

2.5. Kinship Relationship Estimation

Relationships between animals were estimated using kinship-based inference for genome-wide association studies (KING) software v2.2.8 from Manichaikul et al. [7], which applies a robust algorithm to infer pairwise relatedness from genomic data. This approach estimates the degree of relatedness between individuals by measuring IBD from genome-wide SNP data, without requiring complete pedigree information. KING assumes that allele frequencies are known and account for the population structure by incorporating PCA before IBD-based kinship inference.

The relationship coefficient between each pair of individuals, denoted by $\hat{\phi}_{ij}$, represents the probability that two alleles at a locus are identical by descent between individuals (*i* and *j*). In this study, kinship was inferred both within and between family structures, based on heterozygosity and allele sharing across the filtered SNPs [7].

To infer kinship between family relationship structures, the following equation was used:

$$\hat{\phi}_{ij} = \frac{N_{Aa,Aa} - 2N_{AA,\,aa}}{2N_{Aa}^{(i)}} + \frac{1}{2} - \frac{1}{4}\frac{N_{Aa}^{(i)} + N_{Aa}^{(j)}}{N_{Aa}^{(i)}}$$

where $N_{Aa,Aa}$ is the total number of SNPs at which both individuals of the pair are heterozygous, $N_{AA,aa}$ is the total number of SNPs at which both individuals are homozygous on both major and minor alleles, and $N_{Aa}^{(i)}$ and $N_{Aa}^{(j)}$ is the total number of heterozygotes for the *i*-th and *j*-th individuals of the pair.

To infer kinship within family relationship structures, the following equation was used:

$$\hat{\phi}_{ij} = \frac{N_{Aa,Aa} - 2N_{AA,aa}}{N_{Aa}^{(i)} + N_{Aa}^{(j)}}$$

where $\hat{\phi}_{ij}$ is the relationship coefficient of individuals i and j. $N_{Aa,Aa}$ is the total number of SNPs at which both individuals of the pair are heterozygous; and $N_{AA,aa}$ is the total number of SNPs at which one individual is the homozygous major and the other the homozygous minor genotype. $N_{Aa}^{(i)}$ and $N_{Aa}^{(j)}$ are the numbers of heterozygotes for the i-th and j-th individuals, respectively, including SNPs with missing genotypes in both individual pairs. In addition, KING also calculates π_0, the probability that two individuals share zero alleles IBD at a given locus. These two measures were used in combination to classify pairwise relationships, as shown in Table 1.

Table 1. Relationship inference based on estimating kinship coefficient $(\hat{\phi})$ and probability of zero IBD sharing (π_0).

Relationship	$\hat{\phi}$	Inference Criteria ($\hat{\phi}$)	π_0	Inference Criteria (π_0)
Monozygotic twins (MZ)	0.500	$>\frac{1}{2^{3/2}}$	0.000	<0.1
Parent–offspring (PO)	0.250	$(\frac{1}{2^{5/2}}; \frac{1}{2^{3/2}})$	0.000	<0.1
Full siblings (FS)	0.250	$(\frac{1}{2^{5/2}}; \frac{1}{2^{3/2}})$	0.250	(0.1, 0.365)
2nd degree (e.g., half-siblings)	0.125	$(\frac{1}{2^{7/2}}; \frac{1}{2^{5/2}})$	0.500	$(0.365, 1-\frac{1}{2^{3/2}})$
3rd degree (e.g., first cousins)	0.0625	$(\frac{1}{2^{9/2}}; \frac{1}{2^{7/2}})$	0.750	$(1-\frac{1}{2^{3/2}}; 1-\frac{1}{2^{5/2}})$
Other (distant relatives)	~0.000	$<\frac{1}{2^{9/2}}$	1.000	$>1-\frac{1}{2^{5/2}}$

This classification scheme was used to infer kinship within and between individuals of the Nguni and Bonsmara breeds. The threshold values of $\hat{\phi}_{ij}$, defined as the probability that two alleles sampled at random from two individuals, are identical by descend, and π_0, π_1, π_2 denote the probability that the two individuals share zero, one or two alleles identical by descent, respectively. Table 1 provides a robust framework for categorizing relationships into biologically meaningful degrees of relatedness, such as monozygotic twins, parent–offspring, full siblings, second-degree relatives (e.g., half-siblings), and third-degree relatives (e.g., cousins), as well as unrelated pairs. Kingship coefficient is a function of IBD-sharing statistics with relationship $2\phi_{ij} = \pi_{1ij}/2 + \pi_{2ij}$ Inference criteria presented in Table 1 are derived using the power of 2, with the basis that this is the natural scale of the kinship and zero-IBD sharing statistics.

2.6. Parent-Offspring Relationships

All possible pairings of diploid individuals were analyzed using PLINK [1] for computing IBD sharing probabilities, using the 'PI_HAT' parameter. The expected value for most-like first-degree relatedness, which is parent–offspring, is 0.5 for both IBD and genomic relationship (GR). The probabilities of IBD1 (sharing1 allele) and IBD2 (sharing both alleles) across the genome were measured. The proportion of IBDs shared between genomes was calculated as $\mathrm{Pr}_{(\mathrm{IBD1})} \times 0.5 + \mathrm{Pr}_{(\mathrm{IBD2})}$, whereas the GR was calculated using the following equation [13]:

$$G_{ij} = \frac{1}{N}\sum_{n=1}^{N}\frac{(x_{in} - 2p_n)(x_{jn} - 2p_n)}{2p_n(1-p_n)}$$

where G_{ij} is the estimated GR between the i-th and j-th individuals, and x_{in} and x_{jn} refer to the number of reference alleles for the i-th and j-th individuals at SNP_n, respectively, and take a value of 0, 1 or 2 if the genotype at SNP n is *aa*, *Aa* or *AA* (alleles are arbitrarily called A or a), respectively. And p_n is the frequency of the reference allele at SNP n, and N is the number of SNPs. A total of 7021 (Nguni) and 48,205 (Bonsmara) pairings with a PI_HAT value greater than 0.45 were selected before estimating the number of

Mendelian errors (ME) based on a hypothesis that the two individuals were parent and offspring: for example, if an individual had an *AA* SNP genotype at a given locus and the other individual of the pairing had a BB SNP at the same locus, this was considered a Mendelian inheritance error for a parent–offspring relationship. Any pairings showing fewer than 600 ME when using the initial set of SNPs were considered as high-confidence parent–offspring relations (duos).

2.7. Inbreeding Coefficient

Using the PLINK software, the following genetic parameters were determined: inbreeding coefficient (F_{IS}) using het function, and observed and expected heterozygosity (H_O; H_E) using hardy function. The expected heterozygosity (H_E) of a population refers to the probability that any individual in the population is heterozygous at any point, while the observed heterozygosity (H_O) refers to the proportion of the individuals in the population who are heterozygous at a certain locus [18].

2.8. Genetic Distance

Genetic distance between population pairs (pairwise F_{ST}) and the analysis of the main components of variability PCA were performed. Based on the calculated values of genetic distances between individuals per breed, using the SplitsTree v6.4.17 program [19], a phylogenetic tree was created by the neighbor-joining (NJ) method. The normality of the data distribution was tested using the Shapiro–Wilk test, and the Pearson correlation coefficient was estimated using data obtained from the full set of SNP markers between IBD and genomic distance. Genetic distance trees for the breeds were plotted using the SplitsTree program. Furthermore, the number of ancestral populations (*K*) for the two cattle breeds were determined using the cross-entropy approach. The method involved applying *K* values from 1 to 10 to identify the minimum cross-validation error.

3. Results

3.1. Kinship Inference Between Pedigree and Genomic Data

The KING classification of relationships in Bonsmara and Nguni cattle based on pedigree records and genotype data is summarized in Table 2. For Bonsmara, pedigree records indicated distinct relationships, including 75 parent–offspring, zero full siblings, 1132 2nd-degree relatives (e.g., half-siblings), and zero third-degree relatives (e.g., first cousins), with 297 classified as "other" (distant relatives). When genotype-based inference was applied to the same dataset, 62 parent–offspring, 89 full-sibling, 1221 2nd-degree (half-siblings), 54 third-degree (first cousins), and 78 'other' (distant relatives) relationships were identified.

Table 2. Identify-by-descent inference on different degrees of relatedness for Nguni and Bonsmara using both KING and PLINK programs.

	KING (IBD)				PLINK (IBD)	
Relationship	**Nguni**		**Bonsmara**		**Nguni**	**Bonsmara**
	Ped	Infer	Ped	Infer	Infer (IBD > 0.45)	Infer (IBD > 0.45)
MZ	0	2	0	0	0	0
PO	47	49	75	62	40	11
FS	0	13	0	89	2	18
2nd-degree	12	2317	1132	1221	14	43
3rd-degree	1	243	0	54	45	9
Other	3181	617	297	78	3	181

Abbreviations: Ped = pedigree; Infer = Inference: genotype-based inference; IBD = Identity by descent; MZ = monozygotic twins; PO = parent–offspring; FS = full-sibling; 2nd-degree = half-sibling; 3rd-degree = first cousin; Other = distant relatives.

Genotype-based IBD inference, independent of pedigree records, identified 11 parent–offspring, 18 full-sibling, and 43 2nd-degree relationships in the Bonsmara population. In the Nguni population, where incomplete pedigree data were available, genotype-based KING inference identified relationships including 2 monozygotic twin pairs, 49 parent–offspring, 13 full siblings, and 2317 2nd-degree relatives, 243 third-degree (e.g., first cousins), and 617 classified as 'other' (distant relatives) (Table 2).

3.2. Identification of Parent-Offspring Duos

A total SNP set of 465K was used to compute IBD sharing probabilities for all possible pairings of diploid individuals in Nguni and Bonsmara breed, using the 'PI_HAT' parameter on PLINK. There was a total number of 7021 Nguni and 48,205 Bonsmara individual pairings before selecting for IBD paring (distribution of IBD values per breed, Figure 1). After eliminating pairs with PI_HAT of less than 0.45, a subset of 103 Nguni and 262 Bonsmara pairs remained (Table 2). The inferred relationship between KING and PLINK did not provide similar numbers of individuals in each category. It was observed that inferred relationships derived from the KING program were more than those derived from the PLINK program. For PO pairs, we observed 49 relationships compared to 40 PO pairs in Nguni using both KING and PLINK, respectively. We did not observe monozygotic twins with the PLINK program as compared to the KING program. The inferred PO pairs observed for Bonsmara were 62 using KING program, while with PLINK only 11 PO pairs were observed. The distribution of Mendelian error (ME) ranged from 158 to 607 in Nguni and 136 to 156,961 for Bonsmara (Figure 2). Below the ME gap of less than 600 ME [14], a total of 41 parent–offspring pairs for Nguni and 21 offspring pair for Bonsmara duos were obtained (Figure 2).

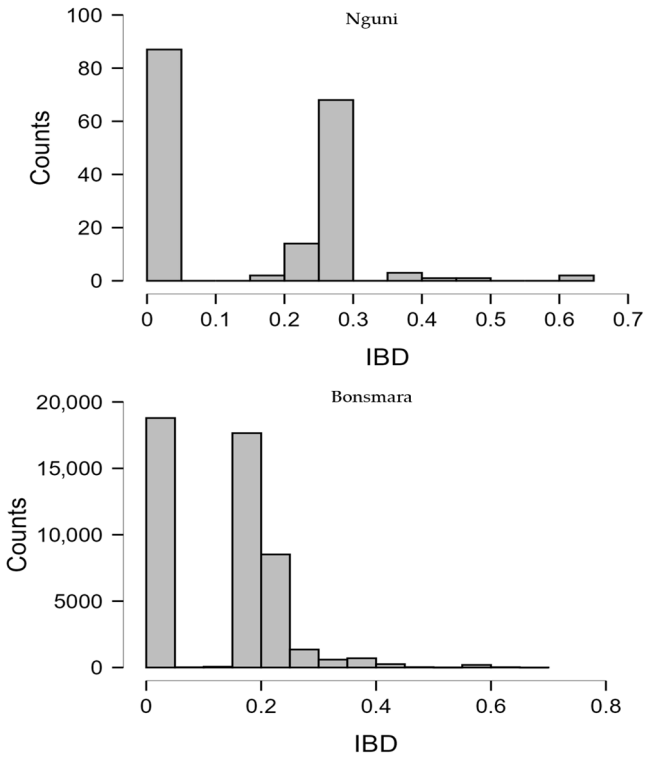

Figure 1. Distribution of IBD between parent–offspring pairs for Nguni (Ngu) and Bonsmara (Bons) breeds. The count of pairs below 0.45 was eliminated to retain pairs with IBD values that can confirm parent–offspring relationships for both breeds.

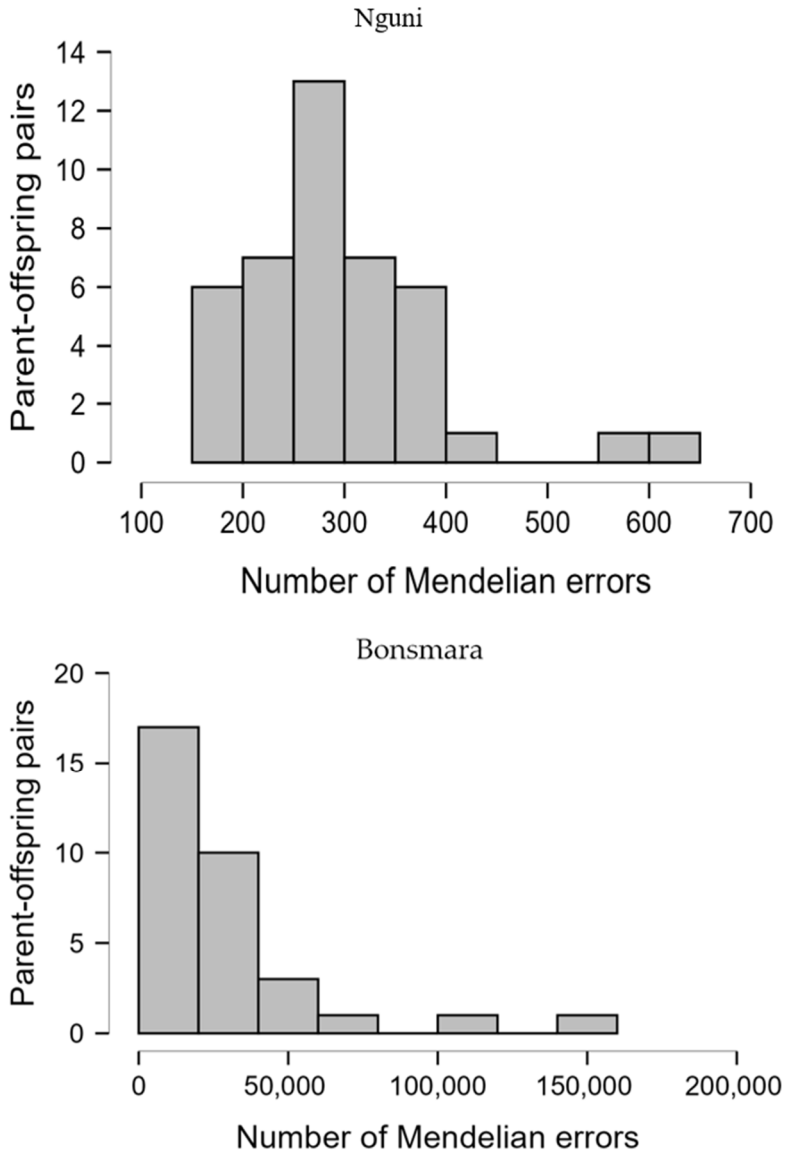

Figure 2. Mendelian error distribution for counts of parent-offspring pairs in Nguni and Bonsmara breeds.

3.3. The Relationship of IBD Versus GR on Parent–Offspring Pairs

Considering the variations in the distribution of GR and IBD values, the correlation between these two metrics was further estimated. The results revealed correlation coefficients of r = 0.87 for Nguni parent–offspring pairs and r = 0.70 for Bonsmara (Figure 3).

3.4. Network of Relationships and Phylogenetic Analysis

A network analysis revealed that animals with a second-degree relationship (half-siblings) were predominant, forming two distinct clusters within both Nguni and Bonsmara breeds. In the Nguni breed, a network showed a densely clustered set of relationships, including parent–offspring pairs, with a few outlier individuals not connected to the main cluster, while in the Bonsmara breed, a first-cousin relationship was observed with an outlier cluster of individuals (Figure S1).

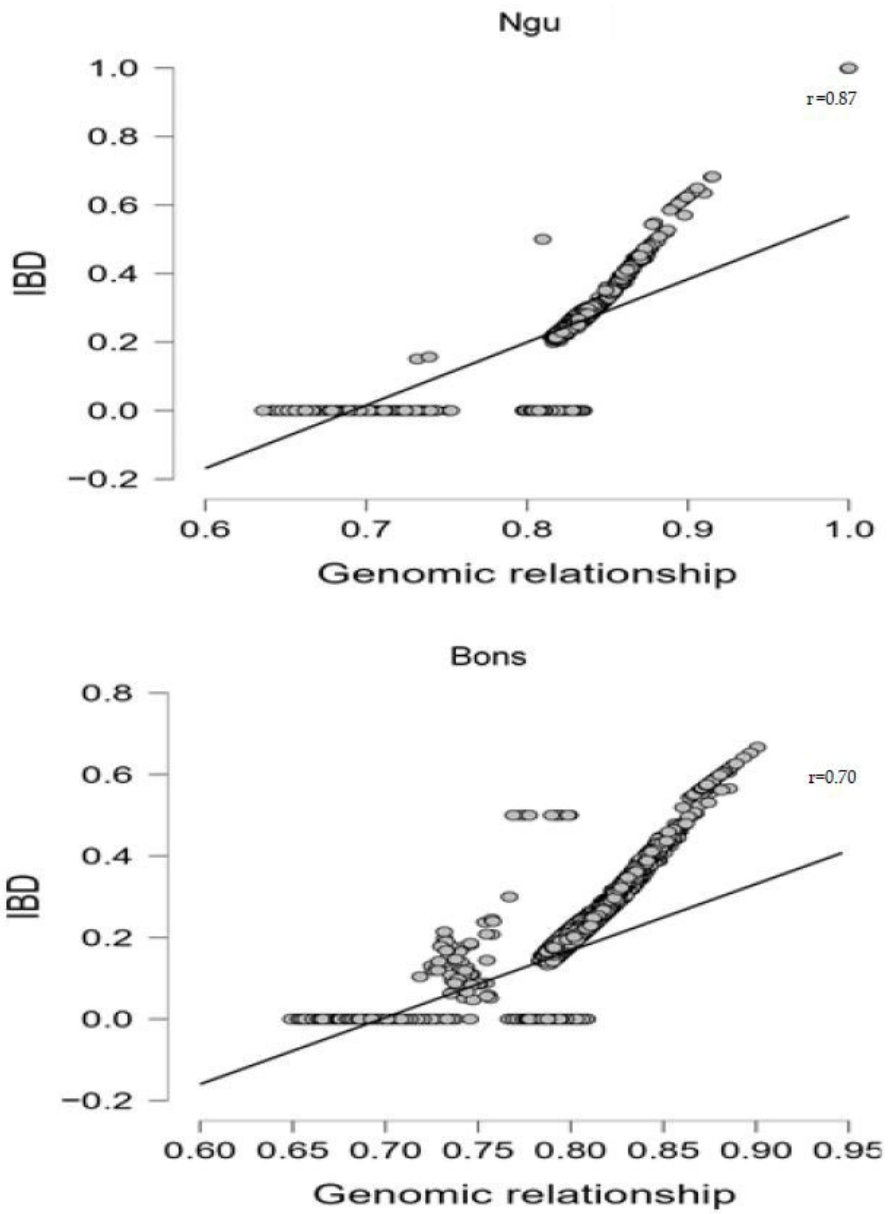

Figure 3. Correlation between IBD and GR. The scatterplot represents parent–offspring pairs, and the line represents relationships.

Phylogenetic analysis in Figure 4 revealed the clustering of breeds based on genetic relationships. The Nguni breed showed segregation into a number of clades with each sub-cluster while we observed few overlapping individuals from all three major clusters, making its own unique sub-cluster. Bonsmara showed a number of individuals connected to groups that differed from the recorded pedigree, suggesting a complex relatedness that was previously not known.

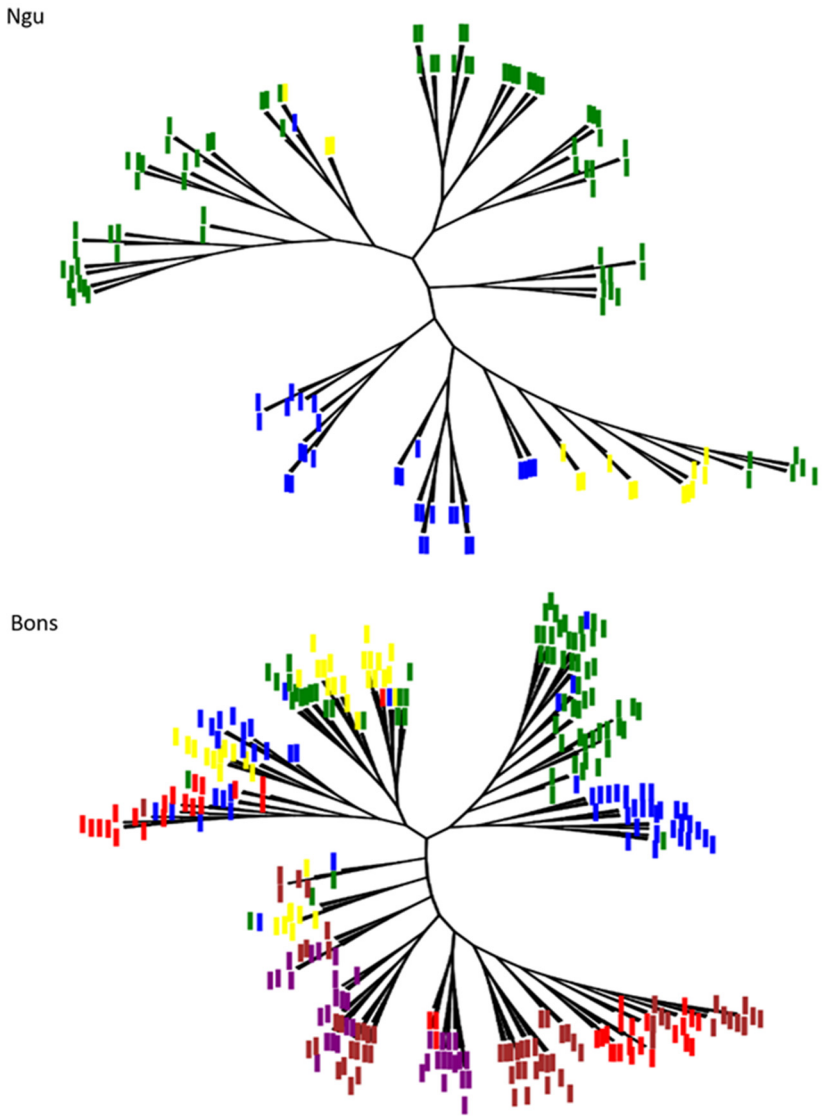

Figure 4. Visualization of genetic variability within Nguni (Ngu) and Bonsmara (Bons) using Neighbor-Joining Network based on genomic relationship matrix. Color code represents family groups within each population. For Nguni breed, (color blue = Ngu1, green = Ngu2 and yellow = Ngu3). And for Bonsmara, (color blue = Bons1, green = Bons2, yellow = Bons3, red = Bons4, brown = Bons5, purple = Bons6).

3.5. Clustering of Family Members Using PCA

PCA was employed to investigate the genetic structure and assess the genetic diversity among Nguni and Bonsmara animals. The PCA scatter plot illustrated the distribution of animals along the first two principal components, which cumulatively captured segregating groups within Nguni and Bonsmara individuals. The first principal component (PCA1) was plotted along the horizontal axis, whereas the second (PCA2) was plotted along the vertical axis. The clustering patterns observed in the PCA plot revealed some overlap of few individuals with clear dominating individuals of a similar cluster/group (Figure 5). The analysis showed clear clustering patterns within Nguni and Bonsmara breeds, with each breed showing three distinct clusters, indicating generational separation between grandparents, parents and offspring. Individuals from the same families clustered closely together, reflecting pedigree-based relatedness (Figure 5).

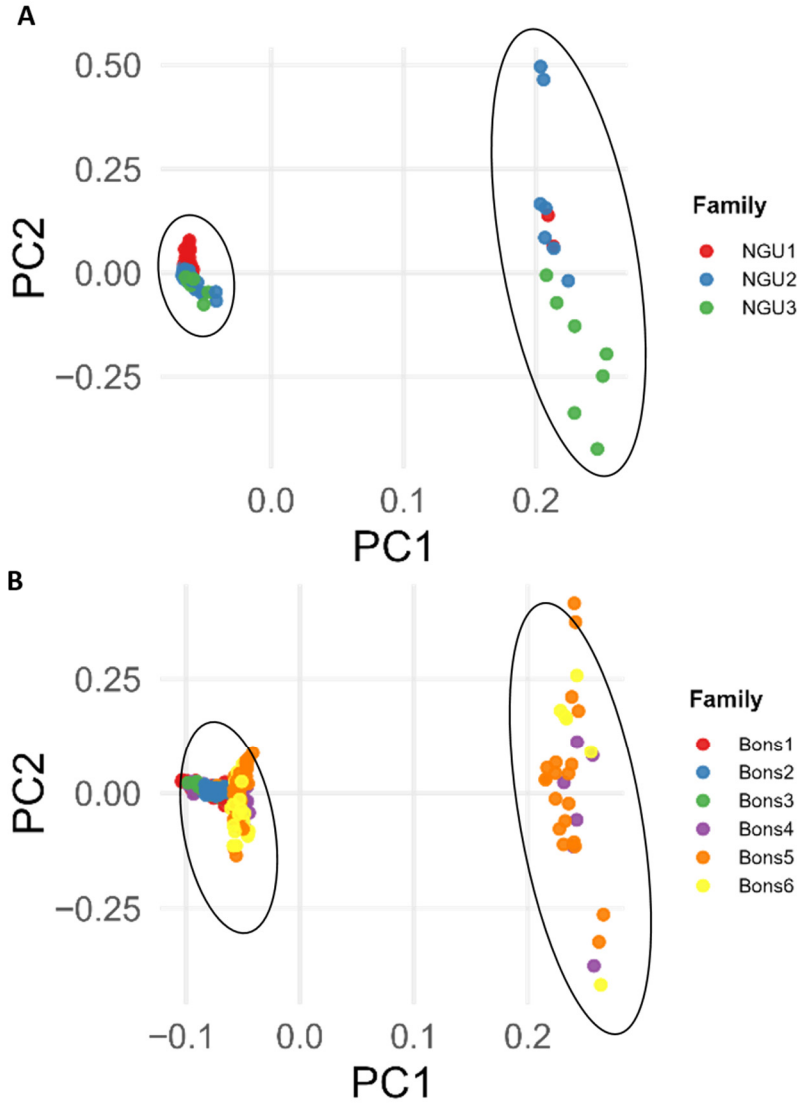

Figure 5. Principal component analysis of genetic variability within Nguni (**A**) and Bonsmara (**B**) populations using more than 400K marker set. The principal components of (PC1) and (PC2) performed in different family groups are shown in different color coordinations.

Table 3 shows the values of observed heterozygosity (H_O) and expected heterozygosity (H_E) and inbreeding coefficient (F_{IS}) estimated from more than 400K marker set. The inbreeding coefficient of Nguni and Bonsmara was 0.13 and 0.10, while genomic relationship was 0.80 and 0.78, respectively.

Table 3. The observed and expected heterozygosity, inbreeding coefficient and genetic distance estimated in Nguni and Bonsmara populations.

Breed	n	H_O	H_E	F_{IS}	GD
Nguni	119	0.25	0.27	0.13	0.80
Bonsmara	311	0.29	0.29	0.10	0.78

Abbreviations: n, number of individuals; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, inbreeding coefficient; GD, genomic distance.

The analysis of genetic structure revealed that the most possible ancestral groups were $K = 5$ and $K = 9$ with the most appropriate number of ancestral populations for Nguni and Bonsmara, respectively (Figure 6).

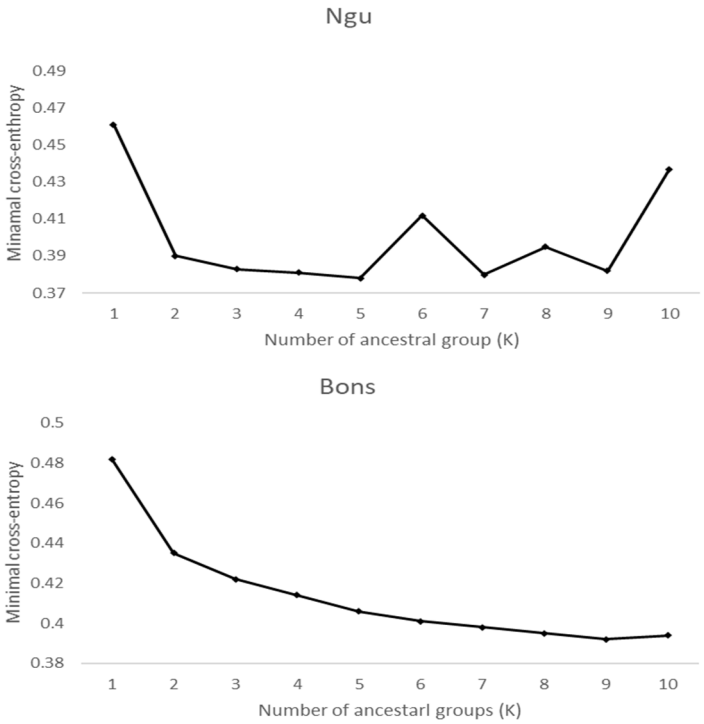

Figure 6. Cross-validation for ancestral groups that have contributed to Nguni and Bonsmara breeds.

4. Discussion

The present study provides new insights into the utility of SNP-based pedigree reconstruction and kinship inference in Sanga cattle breeds. By integrating genomic and pedigree information, this study demonstrated the potential of high-density SNP data to reveal hidden or misclassified relationships, validate parentage, and enhance the reliability of pedigree records. The findings highlight the importance of genomic tools in enhancing management of genetic diversity in populations where pedigree data could be prone to recording errors. The mapping of ambiguous regions and coordinate shifts from LiftOver were reduced by the use of the updated species-specific chain file, ensuring reproducibility of the imputed genotypes from SNP data. All variants that failed to map or mapped to multiple locations were removed from the analysis. The availability of SNP markers, combined with imputation methods and an updated genome assembly for accurate marker positions, enhances the accuracy and reliability of molecular markers for QTL mapping, GWAS studies and pedigree analysis [20]. Understanding genetic structures and inter-individual relationships is critical for designing effective breeding programs and conservation of genetic diversity in livestock populations [10].

This study utilized PCA as an exploratory tool to visualize population structure, as well as SNP-based IBD estimation, using KING and PLINK software to assess the accuracy of recorded pedigrees and infer kinship relationships in two South African cattle breeds, Bonsmara and Nguni. The findings underscore three major insights: pedigree inaccuracies, distinct patterns of population structure suggestive of varied inbreeding levels, and the potential of genomic data to approximate kinship in the absence of reliable pedigree records.

The inference of breed-specific levels of relatedness using IBD values from KING and PLINK software revealed pedigree inconsistencies in the dataset analyzed for Bonsmara. The Nguni breed lacked formal pedigree records, and the genomic data enabled the reconstruction of putative kinship structures for the animals in this study. In Nguni, genomic data revealed 49 and 40 parent–offspring pairs using KING and PLINK, respectively. In contrast, the Bonsmara breed, which had documented pedigrees, showed numerous inconsistencies when pedigree records were compared with SNP-derived kinship estimates. Notably, KING

identified 62 parent–offspring pairs, while PLINK only identified 11. These discrepancies suggest possible underreporting or misclassifications in the pedigree database.

SNP-based IBD inference represents a statistical estimation rather than biological validation [18,21]. The kinship categories generated by KING and PLINK are sensitive to factors such as genotyping quality, marker density, allele frequency distribution, and algorithmic thresholds. Accordingly, while these results strongly suggest mis-recordings, the inferred kinship relationships should be interpreted as probabilistic estimates rather than definitive familial classifications. Our findings are consistent with prior studies, including that of Gorbach et al. [22] and Siderits et al. [21], which documented the consequences of inaccurate pedigree recording on estimates of genetic parameters and selection decisions in both African and European livestock populations.

The identification of two monozygotic twin pairs, which is an uncommon occurrence in cattle, demonstrates the sensitivity of high-density SNP data for detecting close genetic relationships. This could, however, also indicate some potential laboratory errors during processing or duplicated samples. Similar with Bonsmara results, these results should be treated as statistical approximations rather than biological certainties. Nevertheless, in data-sparse systems, SNP-based inference offers a practical tool for mapping genetic relationships and understanding herd demography in the absence of traditional recording materials [18].

In addition to IBD analysis, Mendelian error was used as another robust approach to exclude inconsistent relationships. With the threshold of 600 ME, a total of 41 and 21 parent–offspring pairs were identified for Nguni and Bonsmara, respectively. This ME threshold was similar to Muranty et al. [14] in investigating complex pedigree with generational gaps in grapes. Using Mendelian errors as an exclusion approach to assess relatedness provides a simpler and more efficient method to build pedigrees using SNP markers [23]. This provides a more robust tool that can be considered in breeding strategies that require accurate pedigree information. Moreover, the observed genomic relatedness coefficients exhibited a strong positive correlation with expected IBD values, particularly supporting the identification of parent–offspring pairs. Non-parent–offspring pairs were characterized by genomic relatedness values consistent with an absence of recent IBD, allowing for accurate differentiation within the breed.

The genomic network analysis showed all inferred level of relatedness, where half-sibs was dominant in both breeds. The networks demonstrated the potential to trace interconnectedness of an individual within a population, which could be informative for forensic applications and breeding strategies aimed at reducing inbreeding [19]. However, the network analysis based on neighbor-joining method using genomic relationships confirmed the existence of parent–offspring pairs with a few individuals showing multiple contributions across the two family groups. This could be due to sale or exchange of animals that were subsequently recorded in new family lines, while genetically it originated from different families. Such occurrences are common in the farming industry as they facilitate flow of genetic materials between farms.

Population structure using PCA revealed marked differences between the two breeds. In Bonsmara, individual clustering was tighter, which is consistent with elevated within-population relatedness. For this study, Bonsmara animals were selected from the two regions and from a relatively small dataset. This pattern is likely due to the repeated use of selected elite sires in commercial breeding programs, which can reduce genetic variation. In contrast, the Nguni population showed more dispersed clustering patterns, indicating greater heterogeneity and lower relatedness. These patterns likely reflect the communal breeding practices that dominate indigenous cattle systems, where mating is more random and gene flow occurs across herds through shared grazing and multi-sire

exposure. However, it is essential to note that PCA is fundamentally a dimensionality reduction technique that is designed to summarize genetic variation, but it cannot infer specific pairwise relationships. Clustering patterns observed in PCA may reflect subpopulation structure, historical admixture, or batch effects, but they do not confirm familial relationships. Thus, in this study, PCA was used solely for visualizing population-level genetic structure and not as evidence of individual-level kinship.

The practical value of SNP-based kinship estimation is evident in both breeding contexts. In structured commercial herds like Bonsmara, correcting pedigree errors with genomic data can enhance selection accuracy and reduce inbreeding depression. Incorporating validated relationship matrices into genomic evaluations facilitates the accurate estimation of genetic parameters such as heritability and supports the implementation of genomic estimated breeding values (GEBVs). Strategies like optimum contribution selection (OCS), which balances genetic gain with inbreeding control, depend on reliable kinship matrices which high-density SNP data can provide [24].

In communal systems, where recording of pedigree is not done routinely, genomic tools offer an opportunity for designing structured community-based breeding programs. Reconstructed pedigrees enable the identification of distinct maternal and paternal lines, which can be used to implement rotational mating schemes that preserve genetic diversity. This approach is particularly valuable in the face of environmental stressors, including climate change and emerging diseases, where genetic variability buffers against population collapse [25,26].

Moreover, the real-time integration of genomic data into mobile-based herd management systems, as explored in recent digital agriculture initiatives in sub-Saharan Africa [26,27], could democratize access to breeding insights. By enabling communal farmers to make data-driven mating decisions, these technologies offer scalable pathways to improving livestock productivity while safeguarding genetic diversity.

Overall, this study demonstrates that high-density SNP genotyping can serve as a robust, scalable approach to verifying pedigree accuracy and estimating kinship in South African cattle populations. Whether in formal commercial systems or data-deficient communal herds, genomic tools offer a scientifically grounded framework for enhancing genetic management, supporting sustainable breeding strategies, and securing long-term livestock productivity.

5. Conclusions

This study demonstrated the effectiveness of high-density SNP genotyping in validating and reconstructing kinship relationships within local cattle populations such as Nguni and Bonsmara. The application of genomic data provides a practical and scalable solution for improving pedigree accuracy, enabling more accurate selection decisions and reducing the risk of unintentional inbreeding. This has direct implications for indigenous and locally adapted breeds such as Nguni cattle, where maintaining genetic diversity is essential for long-term resilience and adaptability. Future studies should focus on the cross-generational validation of inferred pedigrees to assess long-term consistency and integrate genomic data into national animal recording systems.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture16060696/s1, Figure S1: Genomic relatedness cluster using KING program for Nguni and Bonsmara breed. Relationship is connected by color coded line showing type of relationship from twinz (MZ) = purple, parent offspring (PO) = red, full siblings (FS) = green, second degree (2nd) = blue and third degree (3rd) = yellow for Nguni and Bonsmara breed.

Author Contributions: Execution, methodology and data collection, investigation, data curation, visualization, data analysis, writing original draft and editing manuscript, K.S.K.; reviewing and editing, N.A.M.; Conceptualization, supervision, resource acquisition, project administration, reviewing of manuscript, A.A.Z.; review and editing of manuscript, supervision, A.M. and E.v.M.-K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Foundation (NRF)—Thuthuka and Red Meat Research and Development South Africa (RMRD—SA).

Institutional Review Board Statement: This study was performed following the guidelines of the Animal Ethics Committee of Agricultural Research Council (AP AEC 2020/17, approved on 25 January 2021) and the University of Pretoria Ethics Committee (NAS374/2022, approved on 3 August 2022). The study commenced once ethical approval was granted and was conducted in accordance with the Animal Welfare Act, which governs the ethical use and protection of animals. However, no physical interaction with animals was conducted in this study; we only used hair samples that were stored internally.

Data Availability Statement: The data presented in this study are not readily available because data is part of ongoing research, and genotypes are generated from commercial animals which raises commercial sensitivity.

Acknowledgments: The authors would like to acknowledge the employees and postgraduate students of the Agricultural Research Council—Animal Production, Animal Genetics Unit, for their technical support during this study, more especially H. Swart who assisted in providing the pedigree records for the study animals.

Conflicts of Interest: The authors declare that there are no conflicts of interest. The funders had no role in the design of the study; collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

Bons	Bonsmara
Ngu	Nguni
SNPs	Single-Nucleotide Polymorphism
ARC—UCSD	Academic Resource Center—University of California, San Diego
ARC	Agricultural Research Council
INTERGIS	Integrated Registration and Genetic Information System
AP	Animal Production
NAS	Natural and Agricultural Sciences
APAEC	Animal Production-Animal Ethics Committee
CA	California
USA	United State of America
DNA	Deoxyribose Nucleic Acid
HD	High-Density
CR	Call Rate
IBD	Identity-By-Descent
PCA	Principal Component Analysis

References

1. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
2. Mapholi, N.O.; Banga, C.; Dzama, K.; Matika, O.; Riggio, V.; Nyangiwe, N.; Maiwashe, A. Prevalence and tick loads in Nguni cattle reared in different environmental conditions across four provinces of South Africa. *Vet. World* **2022**, *15*, 1943–1953. [CrossRef]
3. Kooverjee, B.B.; Soma, P.; van der Nest, M.A.; Scholtz, M.M.; Neser, F.W. Copy number variation discovery in South African Nguni-Sired and Bonsmara-Sired crossbred cattle. *Animals* **2023**, *13*, 2513. [CrossRef]
4. Kunene, L.M.; Muchadeyi, F.C.; Hadebe, K.; Mészáros, G.; Sölkner, J.; Dugmore, T.; Dzomba, E.F. Genetics of base coat colour variations and coat colour-patterns of the south african nguni cattle investigated using high-density snp genotypes. *Front. Genet.* **2022**, *13*, 832702. [CrossRef] [PubMed]
5. Wang, M.; Dzama, K.; Rees, J.; Hefer, C.; Muchadeyi, F. Screening and characterization of copy number cariation in South African nguni cattle using next-generation sequencing data. *J. Anim. Sci.* **2016**, *94*, 61. [CrossRef]
6. Makina, S.O.; Muchadeyi, F.C.; van Marle-Köster, E.; MacNeil, M.D.; Maiwashe, A. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front. Genet.* **2014**, *5*, 333. [CrossRef] [PubMed]
7. Manichaikul, A.; Mychaleckyj, J.C.; Rich, S.S.; Daly, K.; Sale, M.; Chen, W.-M. Robust relationship inference in genome-wide association studies. *Bioinformatics* **2010**, *26*, 2867–2873. [CrossRef]
8. Ramsay, K.; Smuts, M.; Els, H.C. Adding value to South African landrace breeds conservation through utilisation. *Anim. Genet. Resour. Inf.* **2000**, *27*, 9–16. [CrossRef]
9. Pyoos, G.M.; Scholtz, M.M.; MacNeil, M.D.; Theunissen, A.; Neser, F.W.C. Genetic effects from an Afrikaner, Bonsmara, and Nguni three-breed diallel and top-crosses of Angus and Simmental sires. *S. Afr. J. Anim. Sci.* **2020**, *50*, 366–377. [CrossRef]
10. Sanarana, Y.P.; Maiwashe, A.; Berry, D.P.; Banga, C.; van Marle-Koster, E. Evaluation of the International Society for Animal Genetics bovine single nucleotide polymorphism parentage panel in South African Bonsmara and Drakensberger cattle. *Trop. Anim. Health Prod.* **2020**, *53*, 32. [CrossRef]
11. Felius, M.; Beerling, M.-L.; Buchanan, D.; Theunissen, B.; Koolmees, P.; Lenstra, J. On the History of Cattle Genetic Resources. *Diversity* **2014**, *6*, 705–750. [CrossRef]
12. Lashmar, S.F.; Berry, D.P.; Pierneef, R.; Muchadeyi, F.C.; Visser, C. Assessing single-nucleotide polymorphism selection methods for the development of a low-density panel optimized for imputation in South African Drakensberger beef cattle. *J. Anim. Sci.* **2021**, *99*, skab118. [CrossRef]
13. Zhou, Y.; Wang, Q.; Wang, Q.; Yan, Y.; Li, G.; Wu, G.; Yang, N.; Wen, C. Pedigree reconstruction based on genotype data in chickens. *Poult. Sci.* **2024**, *103*, 104327. [CrossRef]
14. Muranty, H.; Denance, C.; Feugey, L.; Crepin, J.L.; Barbier, Y.; Tartarini, S.; Ordidge, M.; Troggio, M.; Lateur, M.; Nybom, H.; et al. Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biol.* **2020**, *20*, 2. [CrossRef]
15. Long, J. Parentage analysis using genome-wide high-density SNP microarray. *Gene* **2021**, *785*, 145605. [CrossRef]
16. Illumina, Inc. *Infinium HD Ultra Assay Guide*; Illumina, Inc.: San Diego, CA, USA, 2009; Volume 11328087, p. 244.
17. Browning, B.L.; Zhou, Y.; Browning, S.R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **2018**, *103*, 338–348. [CrossRef]
18. Jasielczuk, I.; Gurgul, A.; Szmatola, T.; Radko, A.; Majewska, A.; Sosin, E.; Litwinczuk, Z.; Rubis, D.; Zabek, T. The use of SNP markers for cattle breed identification. *J. Appl. Genet.* **2024**, *65*, 575–589. [CrossRef]
19. Huson, D.H.; Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **2006**, *23*, 254–267. [CrossRef]
20. Bowling, A.T. Historical development and application of molecular genetic tests for horse identification and parentage control. *Livest. Prod. Sci.* **2001**, *72*, 111–116. [CrossRef]
21. Siderits, M.; Baumung, R.; Fuerst-Waltl, B. Pedigree analysis in the German Paint Horse: Genetic variability and the influence of pedigree quality. *Livest. Sci.* **2013**, *151*, 152–157. [CrossRef]
22. Gorbach, D.M.; Makgahlela, M.L.; Reecy, J.M.; Kemp, S.J.; Baltenweck, I.; Ouma, R.; Mwai, O.; Marshall, K.; Murdoch, B.; Moore, S.; et al. Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. *J. Anim. Breed. Genet.* **2010**, *127*, 348–351. [CrossRef] [PubMed]
23. Usha, A.P.; Simpson, S.P.; Williams, J.L. Probability of random sire exclusion using microsatellite markers for parentage verification. *Anim. Genet.* **1995**, *26*, 155–161. [CrossRef] [PubMed]
24. Meuwissen, T.H.E.; Odegard, J.; Andersen-Ranberg, I.; Grindflek, E. On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. *Genet. Sel. Evol.* **2014**, *46*, 49. [CrossRef]

25. Food and Agriculture Organization of the United Nations (FAO). *Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*; Scherf, B.D., Pilling, D., Eds.; Food and Agriculture Organization of the United Nations (FAO): Rome, Italy, 2015.
26. Mrode, R.; Ekine-Dzivenu, C.; Gebreyohanes, G.; Ojango, J.; Mwai, O. The Application of Digital-Genomic Tools to Improve Productivity and Profitability in Smallholder Dairy Systems in sub-Saharan Africa. *Anim. Sci. Cases* **2024**, *16*, ascs20240009. [CrossRef]
27. Aliloo, H.; Mrode, R.; Okeyo, A.M.; Gibson, J.P. Ancestral Haplotype Mapping for GWAS and Detection of Signatures of Selection in Admixed Dairy Cattle of Kenya. *Front. Genet.* **2020**, *11*, 544. [CrossRef] [PubMed]