Comparative Analysis of Romanian and Swiss Bovine Populations Using Whole Genome Sequencing and SNP Microarrays

Alexandru Eugeniu Mizeranschi[1*], Radu Ionel Neamț[1], Ciprian Valentin Mihali[1], Vlad Goilean[1], Mihai Carabaș[2], Daniela Elena Ilie[1]

[1]*Research and Development Station for Bovine - Arad, 310059, Arad, Bodrogului 32, Romania*
[2]*Politehnica University of Bucharest, Splaiul Independenţei nr. 313, Bucharest, Romania*

Abstract
The use of SNP microarrays has gained distinguished attention in recent years and the identification of numerous SNPs has proven valuable for genetic evaluation and selection in farm animals. In the current study we compared several bovine populations from Romania and Switzerland at the level of SNPs. Romanian Brown (N=39) and Romanian Spotted (N=245) cattle were genotyped using the Axiom Bovine BovMDv3 SNP microarray. For the Swiss population, we acquired sequencing data from the NCBI SRA database for 80 individuals from three breeds: Brown Swiss (N=20), Original Braunvieh (N=20) and Simmental (N=40). Sequencing data were processed using the Bcbio-nextgen data analysis pipeline and variants were called based on the UMD3.1 reference genome. Common SNPs found from both microarray and sequencing data were retained and genotypes from all the Romanian and Swiss animals were pooled together, resulting in a combined dataset of 48,291 SNPs for 364 individuals. Pairwise comparisons were assessed on the five subpopulations according to Weir and Cockerham's F_{ST} index. Small genetic differences (F_{ST} < 0.05) were found between the Romanian Brown and Swiss Brown Swiss subpopulations and between Romanian Spotted and Swiss Simmental subpopulations. For all the other pairwise comparisons, F_{ST} values were between 0.05 and 0.1, indicating a moderate level of genetic difference among the corresponding subpopulations. The results of fastSTRUCTURE indicated that the most likely number (K) of subpopulations from the pooled dataset was between 8 and 12. Bar plots for K = 5, 8 and 12 confirmed that Romanian Brown and Swiss Brown Swiss subpopulations were genetically similar. However, they also revealed a surprisingly high level of heterogeneity among the Romanian Spotted individuals. As such, future research is required to zoom in on the genetic make-up and explain the most likely sources of heterogeneity for the Romanian Spotted breed. Current results will facilitate a better understanding of genomic selection and its application for improved breeding programs in Romanian cattle breeds.

Keywords: whole genome sequencing; variant calling; SNP microarray; population structure.

1. Introduction

Technological developments of molecular genetics from the last 30 years enabled the identification of molecular markers with economic importance and implementation of so-called *marker assisted selection* (MAS), which was considered a major breakthrough. This improved the effectiveness of breeding programs through the inclusion of a small number of genetic markers for genetic evaluation, which was advantageous for the selection response [1]. MAS has become a valuable tool in selecting for desirable traits. However, its implementation and use has originally been limited to low numbers of markers and, as a consequence, did not

* Corresponding author: Alexandru Eugeniu Mizeranschi, alex.mizeranschi@gmail.com

provide the expected advantages in breeding programs [2]. Therefore, high-throughput genotyping has been devised to use the information from a large number of molecular markers at a time.

In 2009, cattle became one of the first livestock animals to have a fully mapped genome with a total length of approximately 2.8 billion base pairs [1]. Sequencing of the cattle genome (Bos taurus) was finished in 2009 by the Bovine Genome Sequencing and Analysis Consortium. Since then, the genome assembly passed through different improvements. The early assembly versions were Btau_1.0 and Btau_2.0, followed by Btau_3.1, Btau_4.0, Btau_4.2 (2009), Btau_4.6.1 (2012) and Btau_5.0.1 (2015). At the same time, an alternate assembly was done by the Center for Bioinformatics and Computational Biology at the University of Maryland, UMD2, which was published in 2009 [1]. The assemblies have been improved and upgraded with new versions Btau_5.0.1 and UMD3.1.1, respectively. The most recent bovine reference genome is ARS-UCD1.2, submitted by USDA ARS (Agricultural Research Service) in 2018 [3,4]. The ARS-UCD1.2 has 2,715,853,792 bp and has been developed to replace UMD version 3 [5].

Recent progress in molecular genetics allowed the development of genotyping approaches named single nucleotide polymorphism microarrays. Single nucleotide polymorphisms (SNPs) have gained distinguished attention in recent years, becoming a growingly popular tool for farm animals, including bovine. Several SNP arrays from Affymetrix, Illumina and Neogen/GeneSeek were developed for cattle, including lower-density SNP panels (3K, 7K, 15K, 25K markers), medium density (50K SNPs) and up to high-density SNP panels (150K, 250K, 650K, 800K markers). The availability of high numbers of known SNPs and SNP arrays has resulted in new research opportunities for genetic study, such as genetic disease mapping, genome-wide association studies (GWAS) and genomic selection [6-9].

In the current study we have compared several bovine subpopulations from Romania and Switzerland at the level of common SNPs, in order to facilitate a better understanding of genetic diversity, paving the way for genomic selection and its application for improved breeding programs in Romanian cattle breeds.

2. Materials and methods

A total of 364 animals were used in this study, from two Romanian cattle breeds: Romanian Brown (N=39) and Romanian Spotted (N=245) and three Swiss breeds: Brown Swiss (N=20), Original Braunvieh (N=20) and Simmental (N=40). Romanian cattle were genotyped using the Axiom Bovine BovMDv3 SNP microarray. For the Swiss population, sequencing data were acquired from the NCBI SRA database. A number of 80 DNA sequencing samples belonging to the NCBI SRA study ID ERP020037 and BioProject ID PRJEB18113 were downloaded. The data sets have been submitted to the NCBI SRA database between the years 2016 and 2020 by the Institute of Genetics of the University of Bern, Switzerland. The DNA sequencing was performed as part of a wider effort to study diseases and phenotypes with a genomic basis using whole genome sequencing in multiple Swiss cattle breeds [10,11]. **Table 1** lists the SRA Run ID, BioSample ID and breed for the 80 Bos Taurus samples with sequenced DNA that were included in this study.

All sequencing data were processed using the Bcbio-nextgen pipeline version 1.2.3-336920a6 [12], employing current state-of-the-art tools for each of the analysis steps, as such: read mapping was performed with BWA version 0.7.17 [13]; read duplicates were identified using biobambam version 2.0.87 [14**]**; GATK (the Genome Analysis Toolkit) version 4.1.7.0 [15] was used for calling SNPs and small indels; FastQC version 0.11.8 [16] was responsible for collecting quality control statistics for raw sequenced reads; SAMtools/BCFtools version 1.9 [17] and Qualimap version 2.2.2d [18] were used for post-processing and quality control of mapped reads

and variants. The Bcbio-nextgen pipeline was set up as described in a previous study [19].

A number of 284 samples from Romanian cattle were analyzed with the Affymetrix Axiom Bovine BovMDv3 microarray. Data were acquired in the VCF format from the service provider. SNPs with unknown alleles, multiallelic SNPs, indels, duplicate loci and those from unplaced contigs (i.e. not on chromosomes) were removed using BCFtools.

Table 1: NCBI SRA accession IDs for the Swiss sequencing samples used in this study

SRA Run ID	BioSample ID	Breed	SRA Run ID	BioSample ID	Breed
ERR1746308	SAMEA19309168	Simmental	ERR2561401	SAMEA4644741	Original Braunvieh
ERR1746309	SAMEA19309918	Simmental	ERR2561390	SAMEA4644730	Original Braunvieh
ERR1746310	SAMEA19310668	Simmental	ERR2561394	SAMEA4644734	Original Braunvieh
ERR1747028	SAMEA19852168	Simmental	ERR2561400	SAMEA4644740	Original Braunvieh
ERR1747030	SAMEA19853668	Simmental	ERR2561409	SAMEA4644749	Original Braunvieh
ERR1747037	SAMEA19868668	Simmental	ERR2561410	SAMEA4644750	Original Braunvieh
ERR1747038	SAMEA19869418	Simmental	ERR2561424	SAMEA4644764	Original Braunvieh
ERR1747039	SAMEA19870168	Simmental	ERR2561428	SAMEA4644768	Original Braunvieh
ERR1747045	SAMEA19875418	Simmental	ERR2985353	SAMEA5159767	Original Braunvieh
ERR1747048	SAMEA19877668	Simmental	ERR2985354	SAMEA5159768	Original Braunvieh
ERR1766458	SAMEA33004918	Simmental	ERR2985362	SAMEA5159776	Original Braunvieh
ERR2561404	SAMEA4644744	Simmental	ERR2985375	SAMEA5159789	Original Braunvieh
ERR2561405	SAMEA4644745	Simmental	ERR2985376	SAMEA5159790	Original Braunvieh
ERR2985346	SAMEA5159760	Simmental	ERR2985380	SAMEA5159794	Original Braunvieh
ERR2985351	SAMEA5159765	Simmental	ERR2985381	SAMEA5159795	Original Braunvieh
ERR2985352	SAMEA5159766	Simmental	ERR2985382	SAMEA5159796	Original Braunvieh
ERR2985386	SAMEA5159800	Simmental	ERR3890040	SAMEA6528888	Original Braunvieh
ERR2985387	SAMEA5159801	Simmental	ERR3890041	SAMEA6528889	Original Braunvieh
ERR2985389	SAMEA5159803	Simmental	ERR3890042	SAMEA6528890	Original Braunvieh
ERR2985391	SAMEA5159805	Simmental	ERR3890047	SAMEA6528895	Original Braunvieh
ERR2985392	SAMEA5159806	Simmental	ERR1746320	SAMEA19318918	Brown Swiss
ERR2985393	SAMEA5159807	Simmental	ERR1747026	SAMEA19850668	Brown Swiss
ERR2985397	SAMEA5159811	Simmental	ERR1747032	SAMEA19864918	Brown Swiss
ERR2985398	SAMEA5159812	Simmental	ERR2985355	SAMEA5159769	Brown Swiss
ERR2985409	SAMEA5159823	Simmental	ERR2985356	SAMEA5159770	Brown Swiss
ERR2985410	SAMEA5159824	Simmental	ERR2985357	SAMEA5159771	Brown Swiss
ERR2985411	SAMEA5159825	Simmental	ERR2985358	SAMEA5159772	Brown Swiss
ERR2985413	SAMEA5159827	Simmental	ERR2985359	SAMEA5159773	Brown Swiss
ERR3212429	SAMEA5415499	Simmental	ERR2985360	SAMEA5159774	Brown Swiss
ERR3278188	SAMEA5564710	Simmental	ERR2985361	SAMEA5159775	Brown Swiss
ERR3278189	SAMEA5564711	Simmental	ERR2985363	SAMEA5159777	Brown Swiss
ERR3278191	SAMEA5564713	Simmental	ERR2985368	SAMEA5159782	Brown Swiss
ERR3278192	SAMEA5564714	Simmental	ERR2985369	SAMEA5159783	Brown Swiss
ERR3278195	SAMEA5564717	Simmental	ERR2985370	SAMEA5159784	Brown Swiss
ERR3278196	SAMEA5564718	Simmental	ERR2985371	SAMEA5159785	Brown Swiss
ERR3278199	SAMEA5564721	Simmental	ERR2985372	SAMEA5159786	Brown Swiss
ERR3278200	SAMEA5564722	Simmental	ERR2985377	SAMEA5159791	Brown Swiss
ERR3278202	SAMEA5564724	Simmental	ERR2985378	SAMEA5159792	Brown Swiss
ERR3279807	SAMEA5566449	Simmental	ERR2985383	SAMEA5159797	Brown Swiss
ERR3890062	SAMEA6528910	Simmental	ERR2985385	SAMEA5159799	Brown Swiss

The Axiom Bovine BovMDv3 microarray represents variants according to the UMD3.1.1 bovine reference genome [20]. This is based on the older version UMD3.1, with the only difference between them being that version UMD3.1.1 lacks a number of unplaced contigs that were found to arise from contaminating DNA [20]. We used the reference genome UMD3.1 from Ensembl v.94 for mapping the reads of the 80 Swiss samples. However, the slight difference between the two reference genomes did not affect the results of this study, because only variants

placed on chromosomes were considered. For both sets of variants, obtained from microarray and sequencing data respectively, variant statistics were collected using VCFtools version 0.1.16 [21] and BCFtools. Variants were then annotated for possible effects using the Ensembl Variant Effect Predictor (VEP) version 100.2 [22].

Each set of variants was filtered with VCFtools according to MAF, for a minimum of 0.0001, to eliminate monomorphic loci. The two sets of variants were then merged together, resulting in a combined dataset for all samples (microarray+sequencing) and for only those SNPs appearing in both datasets. Then, Weir and Cockerham's F_{ST} index [23] was calculated from the merged dataset for all the possible pairwise subpopulation comparisons using the VCFtools software package. Finally, population structure analysis was performed using the software tool fastSTRUCTURE version 1.0 [24], installed from Bioconda [25].

3. Results and discussion

SNP statistics

The original microarray dataset contained a number of 61,084 loci genotyped for 284 animals. All of the loci were annotated with Ensembl VEP and SNPs were retained according to the level of impact (high or low) of their predicted consequences. This yielded a number of 771 SNPs with a moderate impact on the genome and 46 SNPs with a high level of potential impact. Variants classified as having a moderate impact were generally nonsynonymous SNPs, while those with a high impact were SNPs that occur in splicing regions of genes or ones that introduce or remove start and stop codons into the coding sequences of genes.

The set of variants identified from the sequencing data contained 25,010,193 loci, including both mono-allelic and multi-allelic SNPs and indels. Among these, 78,686 and 7,021 variants were reported by Ensembl VEP as having moderate and high-ranking impact consequences in the genome, respectively. Moderate impact loci included nonsynonymous (missense) SNPs and inframe insertions and deletions, while high impact variants included those occurring in splicing regions of genes, frameshift insertions or deletions and variants that introduce or remove start and stop codons into the coding regions of genes.

After intersecting the filtered microarray variants and the sequencing variants, 48,291 monoallelic SNPs were found in common between the microarray and sequencing samples. The corresponding genotypes were merged, resulting in a dataset of 48,291 loci and 364 animals. This combined dataset was further used to compute and compare Weir and Cockerham's F_{ST} index for each of the underlying subpopulations.

In the remainder of this article, we use the following breed naming notations: RBR: Romanian Brown; RSP: Romanian Spotted; SBS: Swiss Brown Swiss; SOB: Swiss Original Braunvieh and SSM: Swiss Simmental.

Table 2: mean F_{ST} values for pairwise population comparisons, in ascending order.

Population 1	Population 2	Mean F_{ST}	Difference
RBR	SBS	0.02552	low
RSP	SSM	0.034358	low
SOB	SBS	0.051407	moderate
SSM	SOB	0.059214	moderate
SSM	SBS	0.071569	moderate
RBR	SOB	0.072014	moderate
RSP	SOB	0.078708	moderate
RBR	SSM	0.085368	moderate
RSP	SBS	0.090377	moderate
RBR	RSP	0.09417	moderate

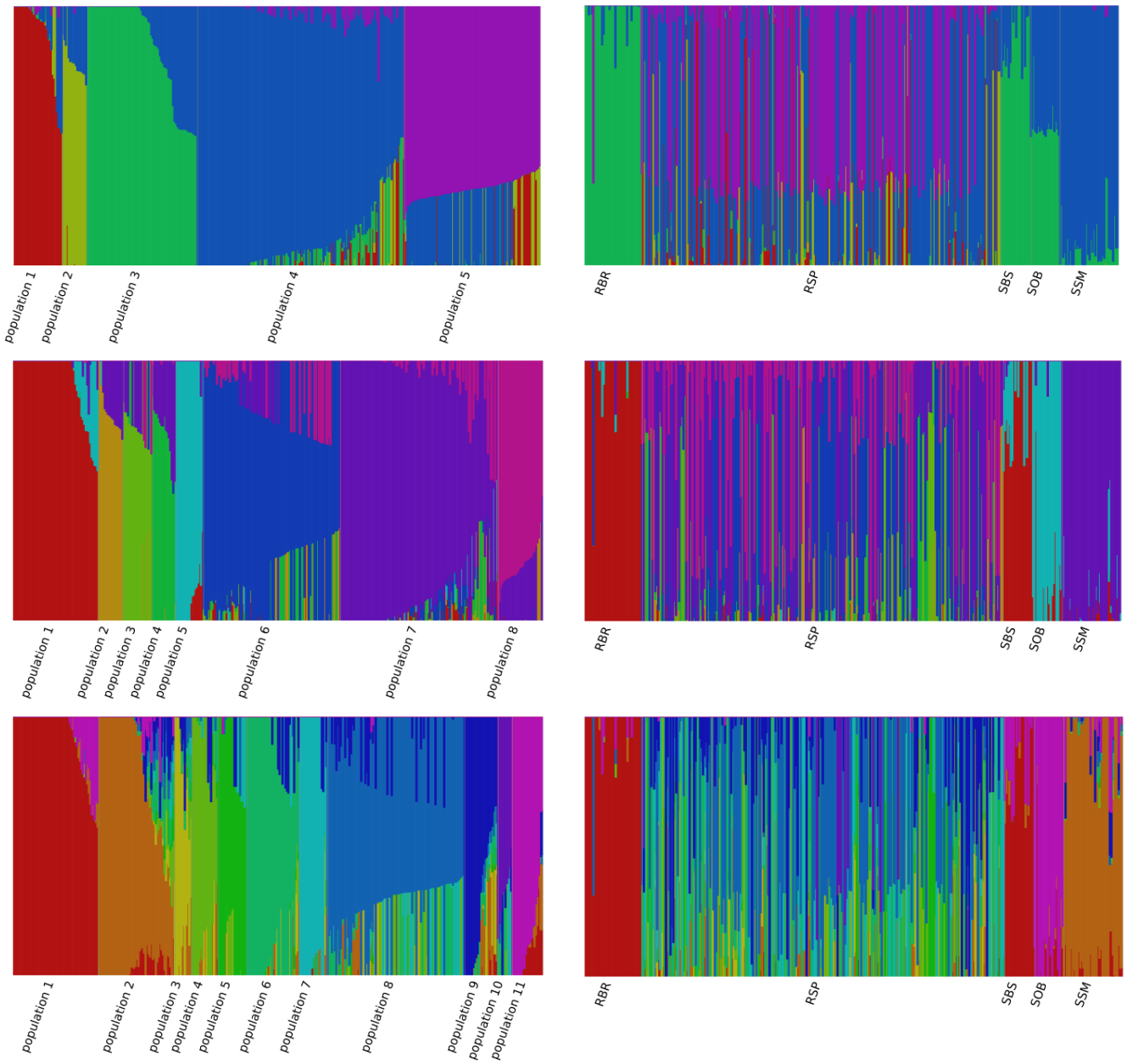

Figure 1: bar plots depicting the results obtained with fastSTRUCTURE for *K* = 5 (top), 8 (middle) and 12 (bottom). The left side shows individuals labeled according to the classes uncovered by fastSTRUCTURE. The right side labels individuals according to their known breeds.

Mean F_{ST} values for population comparisons

A possible interpretation of the fixation index, F_{ST}, is as follows [26]: $F_{ST} < 0.05$ denotes little genetic differences; F_{ST} between 0.05 – 0.15 depicts moderate genetic differences; F_{ST} between 0.15 – 0.25 indicates large genetic differences; finally, $F_{ST} > 0.25$ denotes very large genetic differences.

The obtained pairwise comparison results of the five subpopulations from this study are listed in **Table 2**.

Finally, an investigation of population structure was carried out using fastSTRUCTURE, for *K* ranging between 2 and 20. The utility *chooseK.py* was executed on the results and indicated that the number of subpopulations from the combined dataset was between 8 and 12.

Bar plots from fastSTRUCTURE results for *K* = 5, 8 and 12 confirmed that RBR and SBS subpopulations were genetically similar. However, they also revealed a relatively high level of heterogeneity among the RSP individuals.

Figure 1 shows bar plots for fastSTRUCTURE results for *K* = 5, 8 and 12. The left side shows populations according to fastSTRUCTURE's clustering, while the right side shows individuals labeled according to their actual (known) subpopulation origins (breeds). Note: for K = 12, *population 12* as uncovered by fastSTRUCTURE (bottom-left side in the figure) was comparatively small and, as such, does not appear labeled in the figure. Based on colors used to represent the different populations, several correspondences can be made between the clusters uncovered by fastSTRUCTURE and the different breeds of the animals included in this study. For example, for the top two images showing *K*=5, population 3 (shown predominantly in green) appears to comprise individuals from the RBR, SBS and SOB breeds, whereas the RSP spans individuals clustered in *population 1*, *population 2*, *population 4* and *population 5*, with the SSM breed also corresponding to a subset of individuals from *population 4*.

Similar observations can be made for the plots corresponding to *K*=8 and *K*=12. In both scenarios, *population 1* (shown in red) contains individuals from the RBR and SBS breeds, once again pointing out the genetic similarity of these two breeds. SOB appears more genetically distinct, as its individuals were clustered separately starting from *K*=8 into *population 5*. Finally, the RSP breed shows considerable heterogeneity, where its individuals were clustered into multiple fasSTRUCTURE subpopulations, whereas the SSM breed is relatively homogeneous. For *K*=8, the SSM and RSP subpopulations partially overlap through *population 7*. However, for *K*=12, this overlap appears less pronounced, through *population 2*.

4. Conclusions

We provided an overview of genomic variation in several bovine subpopulations from Romania and Switzerland using 48,291 SNPs found in common in SNP microarray and sequencing data for a total of 364 animals. The results showed small genetic differences between the Romanian Brown and Swiss Brown Swiss subpopulations and between Romanian Spotted and Swiss Simmental subpopulations. However, they also revealed a surprisingly high level of heterogeneity within the Romanian Spotted subpopulation.

Further research with larger sets of animals from multiple herds will be required to reveal additional information about the sources of heterogeneity for the Romanian Spotted breed. Current results will facilitate a better understanding of subpopulation structure for Romanian cattle breeds and support the implementation of genomic selection for improved breeding programs in Romania.

References

1. Zimin A.V., Delcher A.L., Florea L., Kelley D.R., Schatz M.C., Puiu D. A whole genome assembly of the domestic cow, Bos taurus, Genome Biology, 2009, 10, 4: R42, doi: 10.1186/gb-2009-10-4-r42: 10 p.
2. Dekkers J.C.M., Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons, Journal of Animal Science, 2004, 82, 313–328
3. https://bovinegenome.elsiklab.missouri.edu
4. Rosen, B.D., Bickhart, D.M., Schnabel, R.D., Koren, S., Elsik, C.G., Zimin, A., Dreischer, C., Schultheiss, S., Hall, R., Schroeder, S.G., Van Tassell, C.P., Smith, T.P.L., Medrano, J.F., Modernizing the bovine reference genome assembly. Proc. World Congr. Genet. Appl. Livest. Prod., Vol. Mol. Genet. 2018, 3, 802.
5. Null, D.J., Van Raden, P.M., Rosen, B.D., O'Connell, J.R., Bickhart, D.M., Using the ARS-UCD1.2 reference genome in U.S. evaluations. Interbull Bulletin., 2019, 55, 30-34.
6. Boison, S.A., Utsunomiya, A.T.H., Santos, D.J.A., Neves, H.H.R., Carvalheiro, R., Mészáros, G., Utsunomiya, Y.T., do Carmo, A.S., Verneque, R.S., Machado, M.A. and Panetto, J.C.C., Accuracy of genomic predictions in Gyr (Bos indicus) dairy cattle. Journal of Dairy Science, 2017, 100(7), pp.5479-5490.
7. Brown, A., Ojango, J., Gibson, J., Coffey, M., Okeyo, M., Mrode, R., Genomic selection in a crossbred cattle population using data from the dairy genetics East Africa project. J Dairy Sci, 2016, 99, 7308–7312. doi: 10.3168/jds.2016-11083
8. Murgiano, L., Jagannathan, V., Benazzi, C., Bolcato, M., Brunetti, B., Muscatello, L.V., Dittmer, K., Piffer, C., Gentile, A., Drögemüller, C., Deletion in the EVC2 gene causes chondrodysplastic dwarfism in Tyrolean Grey cattle. PloS One, 2014, 9, e94861. doi: 10.1371/journal.pone.0094861
9. Tiezzi F, Parker-Gaddis KL, Cole JB, Clay JS, Maltecca, C. A Genome-Wide Association Study for Clinical Mastitis in First Parity US Holstein Cows Using Single-Step Approach and Genomic Matrix Re-Weighting Procedure, PLoS ONE, 2015, 10(2), e0114919. doi:10.1371/journal.pone.0114919
10. Häfliger, I.M., Hirter, N., Paris, J.M., Wolf Hofstetter, S., Seefried, F.R. and Drögemüller, C., A de novo germline mutation of KIT in a white-spotted

Brown Swiss cow. Animal genetics, 2020, 51(3), pp.449-452.
11. Häfliger, I.M., Sickinger, M., Holsteg, M., Raeder, L.M., Henrich, M., Marquardt, S., Drögemüller, C. and Lühken, G., An IL17RA frameshift variant in a Holstein cattle family with psoriasis-like skin alterations and immunodeficiency. BMC genetics, 2020, 21, pp.1-10.
12. https://github.com/bcbio/bcbio-nextgen
13. Li, H. and Durbin, R., Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 2009, 25(14), pp.1754-1760.
14. Tischler, G. and Leonard, S., Biobambam: tools for read pair collation based algorithms on BAM files. Source Code for Biology and Medicine, 2014, 9(1), p.13.
15. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D. and Shakir, K., Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv, 2017, p.201178.
16. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
17. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., The sequence alignment/map format and SAMtools. Bioinformatics, 2009, 25(16), pp.2078-2079.
18. García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F. and Conesa, A., Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics, 2012, 28(20), pp.2678-2679.
19. Mizeranschi, A.E., Mihali, C.V., Neamț, R.I., Carabaș, M. and Ilie, D.E., September. Distributed bioinformatics analyses on an SGE cluster, for variant calling on bovine whole-genome sequencing samples. In 2019 IEEE International Conference on Advanced Scientific Computing (ICASC) (pp. 1-6). IEEE, 2019
20. Bos Taurus reference genome UMD3.1.1: https://www.ncbi.nlm.nih.gov/assembly/GCF_000003055.6/
21. Merchant, S., Wood, D.E. and Salzberg, S.L., Unexpected cross-species contamination in genome sequencing projects. PeerJ, 2014, 2, p.e675.
22. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G., The variant call format and VCFtools. Bioinformatics, 2011, 27(15), pp.2156-2158.
23. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F., The Ensembl Variant Effect Predictor. Genome biology, 2016, 17(1), p.122.
24. Weir, B.S. and Cockerham, C.C., Estimating F-statistics for the analysis of population structure. Evolution, 1984, pp.1358-1370.
25. Raj, A., Stephens, M. and Pritchard, J.K., fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics, 2014, 197(2), pp.573-589.
26. Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R. and Köster, J., Bioconda: sustainable and comprehensive software distribution for the life sciences. Nature methods, 2018, 15(7), pp.475-476.
26. Hartl, D.L., Clark, A.G. and Clark, A.G., Principles of population genetics, 1997, (Vol. 116). Sunderland, MA: Sinauer associates.