





Deliverable D8.1

On-line tool for product authenticity testing based on SNP data

Due date: 31/01/2026

Lead contractor: USTO

Partners involved: IBBA-CNR, IDELE, INRA-Maroc, ESAM

Dissemination level: PU (public)

Co-funded by the European Union

Introduction

The conservation of livestock genetic diversity is fundamental for ensuring food security, adaptive potential, and resilience in the face of climate change and evolving production systems. Local breeds often harbor unique adaptive traits shaped by centuries of selection under specific environmental and socio-economic conditions. However, the rapid intensification of animal production, the replacement of local breeds by high-yielding commercial ones, and genetic erosion resulting from unplanned crossbreeding have led to a significant decline in livestock biodiversity.

Reliable identification and monitoring of genetic resources are therefore essential for conservation planning and sustainable management. Advances in molecular genetics, particularly through Single Nucleotide Polymorphism (SNP) genotyping, now allow fine-scale characterization of population structure and breed distinctiveness. Yet, the challenge remains to translate this genomic information into practical tools that can be applied in real-time conservation and breeding decisions.

Development of an Online Tool for Local Product Authenticity

Overview and Objectives

As part of Task 8.1, an online computational tool was developed to support the authentication and traceability of local animal-derived products through genomic information. The tool addresses a key need in livestock genomics: the efficient discrimination of individual genotypes and the assignment of individuals to specific populations or breeds using a minimal set of highly informative genetic markers.

The ability to accurately discriminate individual genotypes plays a fundamental role in several applications, including genomics-assisted breeding programs, biodiversity conservation, pedigree verification, traceability along animal-derived food chains, and germplasm management. In the context of product authenticity and traceability, reliable genetic identification is essential to verify the origin of animals and to ensure transparency throughout the production chain.

A major methodological challenge lies in identifying the smallest possible subset of single nucleotide polymorphism (SNP) markers that can uniquely distinguish a large number of individuals. This task is particularly complex when dealing with modern genomic datasets generated through high-density SNP arrays or sequencing technologies. Livestock genomic datasets

typically include hundreds to thousands of animals and tens to hundreds of thousands of SNP markers. Under these conditions, an exhaustive evaluation of all possible SNP combinations becomes computationally infeasible.

The problem is further complicated by the presence of closely related individuals, such as half-siblings or animals belonging to the same flock or herd structure. These genetic relationships increase genomic similarity among individuals and consequently reduce the discrimination power of individual markers. Although standard quality control procedures and linkage disequilibrium pruning strategies are commonly used to reduce redundancy in genomic datasets, these approaches are not specifically designed to optimize individual discrimination and therefore may not yield the most informative marker subsets for identification purposes.

To address these limitations, the project developed a dedicated computational framework specifically designed to maximize genotype discrimination while minimizing the number of markers required. The resulting implementation was integrated into an online tool that allows users to process genomic datasets and identify highly informative SNP panels suitable for individual identification and population assignment.

Methodological Framework

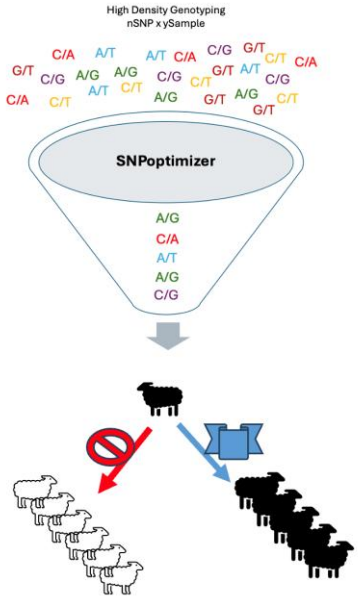
The tool developed by SCALA-MEDI partners provides several key functionalities. First, it enables the integration of genotype data generated from multiple marker platforms, including SNP arrays of different densities as well as sequencing-based datasets. This flexibility allows the tool to be applied across diverse livestock genomic resources and facilitates interoperability among datasets generated using different technologies.

Second, the tool implements algorithms designed to select a minimal subset of highly informative SNP markers that maximize the ability to distinguish individuals within large genomic datasets. By focusing on the most informative markers, the approach significantly reduces the dimensionality of the data while maintaining strong discriminatory power.

Third, the tool supports the assignment of individuals to specific populations or breeds. This functionality is particularly relevant for traceability applications, where verifying the origin of animals or animal-derived products is essential for quality assurance, certification schemes, and the protection of local production systems.

Importantly, the computational framework underlying the tool dramatically improves efficiency compared with exhaustive SNP evaluation approaches. Benchmark analyses indicate that the computational requirements can be reduced by up to fifty-fold while maintaining high accuracy in genotype discrimination. This improvement makes the method suitable for practical use in large-scale genomic datasets typical of livestock research and production systems.

The online implementation of the tool provides a user-friendly interface that allows researchers and stakeholders to upload genotype data, configure analysis parameters, and obtain optimized SNP panels for individual discrimination and population assignment. By combining computational efficiency with practical usability, the tool represents a valuable resource for supporting authenticity verification and traceability in animal-derived food chains.



The development of this tool contributes directly to the objectives of Task 8.1, providing a practical solution for genomic-based authentication of local livestock products and strengthening the capacity to monitor and protect regional production systems.

The SNPoptimizer tool was implemented entirely in the R programming environment (R Core Team, 2024) and consists of two main components. The first component is a preprocessing workflow (main.R), designed to prepare genomic datasets prior to analysis and adaptable to different datasets or experimental designs. The second component is an interactive web interface

developed using the Shiny framework (app.R) (Chang et al., 2025), which allows users to explore data, visualize results, and perform marker optimization in an accessible and reproducible environment.

The preprocessing workflow is responsible for the initial handling of genotype data. This includes importing SNP datasets, performing quality control procedures, encoding genotypes into the required format, and applying marker filtering steps to ensure that only informative loci are retained for downstream analyses. These procedures are designed to accommodate genomic data generated from different platforms commonly used in livestock genomics, including SNP arrays and sequencing-derived markers.

The Shiny-based interface provides a user-friendly environment in which the optimization procedure can be executed interactively. Through this interface, users can run the SNP selection algorithm, monitor intermediate analytical outputs, and visualize the performance of candidate marker sets. The platform ultimately returns an optimized panel of discriminatory SNP markers suitable for distinguishing individual animals or assigning them to specific sheep populations or breeds. This functionality is particularly relevant for applications related to traceability, breed authentication, and genomic characterization of sheep production systems.

The SNPoptimizer package is at the following link: https://nameless8.shinyapps.io/snpoptimizer/.

Conclusion

The development of the SNPoptimizer online tool represents a significant step toward translating genomic data into practical solutions for livestock traceability and product authentication. By combining efficient computational algorithms with an accessible web-based interface, the platform enables the identification of minimal yet highly informative SNP panels capable of accurately discriminating individuals and assigning them to specific populations or breeds. This approach substantially reduces computational burden while maintaining high discriminatory power, making it suitable for large genomic datasets commonly generated in livestock research. Ultimately, the tool provides researchers, breeders, and stakeholders with a practical resource to support genomic-based verification of animal origin, strengthen transparency along animal-derived food chains, and contribute to the conservation and sustainable management of local livestock genetic resources.