POLITECNICO DI TORINO
Repository ISTITUZIONALE

Computer-Assisted Molecular Traceability for Dairy Farming Products

Original
Computer-Assisted Molecular Traceability for Dairy Farming Products / Rossi, F., Modesto, P., DI CARLO, S., Politano, G.M.M., Savino, A., Acutis, P.L., Benso, A.. - ELETTRONICO. - (2016), pp. 1-11. (INTERNATIONAL WORK-CONFERENCE ON BIOINFORMATICS AND BIOMEDICAL ENGINEERING (IWBBIO 2016) Granada, ES 20-22 April, 2016).

Availability:
This version is available at: 11583/2649757 since: 2016-09-16T16:07:53Z

Publisher:

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

03 November 2026

Computer-Assisted Molecular Traceability for Dairy Farming Products

Francesco Rossi[1], Paola Modesto[2], Stefano Di Carlo[1], Gianfranco Politano[1], Alessandro Savino[1], Pier Luigi Acutis[2], and Alfredo Benso[1]

[1] Politecnico di Torino, Control and Comp. Engineering Department, Torino, Italy
E-mail: <firstname>.<lastname>@polito.it
Web Page: http://www.sysbio.polito.it
[2] Istituto Zooprofilattico Sperimentale del Piemonte Liguria e Valle d'Aosta, Torino, Italy
E-mail: <firstname>.<lastname>@izsto.it

Abstract. Food integrity and food safety have received much attention in recent years due to the dramatic increasing number of food frauds. In this article we analyze dairy products for which one of the crucial issues is traditional cheese traceability. In this paper we propose a computer-assisted molecular traceability system able to certify a traditional dairy product. We investigate the use of short tandem repeat analysis data processed by a Covariance Matrix Adaptation Evolution Strategy algorithm in order to predict the traceability between dairy products and the corresponding producer and to highlight possible adulterations and/or inconsistencies. Preliminary results collected from two farms are presented in this study to show the capability of the proposed algorithm in a real setup.

Keywords: Traceability, Dairy Farms Products, Genetic Pool Analysis, STRs, CMA-ES, Food Safety,

1 Introduction

Food integrity and food safety have received much attention in recent years due to the dramatic increasing number of food frauds. Anti-adulteration methods are increasingly required by authorities designated to food controls. Traceability is a useful method to guarantee foodstuff quality and safety, to guarantee hygiene standards, to protect consumer's choices and health and also to value high quality traditional food such as protected designation of origin (PDO) and protected geographical indication (PGI).

Over the past years DNA analysis has been widely recognized to be an effective tool to deal with genetic traceability issues. The DNA constitutes an objective test and an irrefutable proof and can be applied to many food matrices, even when processed. Therefore, DNA analysis is gaining a key role in tracing and testing food origin and safety.

In this article we analyze dairy products for which one of the crucial issues is traditional cheese traceability. In the case of frauds, it may occur that a selected

cheese product that should be produced by a set of certified farms is produced by other farms without authorization. Since the hygienic conditions of the unauthorized farms can not be verified, consequences on consumer's health may occur. Traceability of dairy products through DNA analysis involves some technical challenges since cheese (CH) is produced from bulk milk (BM), which contains DNA from different cows of the farm, and undergoes through consistent changes during ripening.

In this paper we propose a computer-assisted molecular traceability system able to certify a traditional dairy product. We investigate the use of short tandem repeat (STR) analysis to create a DNA fingerprint of small dairy farms and to assign dairy products (milk and cheese) to the corresponding producer. So far, STR analysis has been applied to the blood combining a few individuals and exclusively for genetics population analysis [14, 15, 13], or to the milk in order to identify quantitative trait locus (QTL) associated with traits in animal science [3, 12]. However, the application of STR analysis to traceability of dairy products is a complex issue. Dairy products contain the DNA belonging to several different individuals, preventing the possibility to perform individual traceability. Typically, molecular traceability in the dairy field is only applied to milk origin certification used for cheese production (species and breed certification). These methods enable to trace the milk-breed origin and therefore are useful for the traceability of mono-breed products (e.g., Parmigiano Reggiano cheese derived from *'disolabruna'* red cows). To the best of our knowledge, this work is the first attempt to explore this approach for traceability of food products.

Two farms were included in this study and blood and milk samples of two different breed cows were collected. Then bulk milk and derived cheese were sampled monthly for one year for each farm. DNA was extracted from collected instances and finally a panel of known STR has been tested all over the samples. The obtained STR genetic dataset has been analyzed through a Covariance Matrix Adaptation Evolution Strategy (CMA-ES) algorithm in order to predict the traceability between dairy products and the corresponding producer and to highlight possible adulterations and/or inconsistencies.

Results showed that bulk milk and derived cheese present an STR profile composed of a subgroup of STR among those of all the animals the dairy product originated from, and the profile can be efficiently used to trace the origin of a dairy product.

2 Materials and Methods

In this section, we describe the available STR global dataset collection procedure underlining the raw data manipulation, and we present the proposed Computer-Assisted Molecular Traceability system and its implementation based on the CMA-ES [17] implementation available in R [16].

2.1 STR Dataset

Two farms with different geographic locations have been considered in this study: the fist one is placed in the province of Biella (Italy) and includes 12 *'Pezzata Rossa d'Oropa'* breed cows, the second one is located in the province of Imperia (Italy) and includes 14 *'Bruna Alpina'* breed cows.

At the beginning of the study, appointed veterinaries collected blood and milk samples of each cow, then they monthly sampled bulk milk and derived cheese for 12 months. All collected samples have been cold-stored for the tuning of the analysis protocol and the choice of the best genotyping process.

The main steps of the STR selection and data generation can be summarized as follows:

- **Sample Collection**: DNA was extracted from blood and milk somatic cells with the ReliaPrep^{TM} gDNA Tissue Miniprep System (Promega) and from cheese with the QIAamp DNA Stool Mini Kit (Qiagen);
- **First STRs Selection**: from a panel of 280 available STRs, 57 STRs were selected taking into account:
  - greater allelic richness;
  - homogeneous distribution on the genome and the amplification product length between 90 and 300 base pairs (bp);
  - absence of linkage disequilibrium.
- **Protocol Analysis Tuning**: the investigation consisted of DNA sample extraction, simplex polymerase chain reaction (PCR) protocol test and then a capillary electrophoresis sequencing product analysis comparison with regard to sequences deposited in Genbank®[4] using the Basic Local Alignment Search Tool (BLAST) software [1];
- **Second STRs Selection**: 20 out of the 57 previously selected STRs were chosen by showing amplification products with variable lengths among different items;
- **Final Genotyping Process and Data Extraction**: the analysis of the amplification products was performed by capillary electrophoresis and polymer separation. Following, fragment length determination was achieved using the software Genemapper®(Applied Biosystems) and 500 ROXTM size standard (Applied Biosystems). At this stage electropherogram traces analysis was conducted by a single operator together with the stutter tracks correction. Eventually the allele's frequency attribution for mixed (pool) samples (bulk milks and cheeses) was performed and the relative fluorescence units (RFU) estimation of the alleles measure was computed. We decided to use the peak height of the electropherogram track of each allele as an indication of its quantity.

STRs data organization Once the genotyping process was completed, obtained data have been organized in a tabular format (Table 1) reporting the allele frequencies for each STR and for each cow where:

- n is the number of processed STRs;

- m is the number of cows presented in the examined farm;
- $a^{(m\text{-th},n\text{-th})}$ is the specific allele's dimension (bp) of the m-th cow for the n-th STR with the polymorphism occurrence of being heterozygote ($a^{(..)x} \neq a^{(..)y}$) or homozygote ($a^{(..)x} = a^{(..)y}$).

Cows List	STR1	STR2	STR3	...	STRn
COW1	$a^{(1,1)x}/a^{(1,1)y}$	$a^{(1,2)x}/a^{(1,2)y}$	$a^{(1,3)x}/a^{(1,3)y}$	...	$a^{(1,n)x}/a^{(1,n)y}$
COW2	$a^{(2,1)x}/a^{(2,1)y}$	$a^{(2,2)x}/a^{(2,2)y}$	$a^{(2,3)x}/a^{(2,3)y}$	...	$a^{(2,n)x}/a^{(2,n)y}$
COW3	$a^{(3,1)x}/a^{(3,1)y}$	$a^{(3,2)x}/a^{(3,2)y}$	$a^{(3,3)x}/a^{(3,3)y}$	...	$a^{(3,n)x}/a^{(3,n)y}$
...	...	...	...	...	...
COWm	$a^{(m,1)x}/a^{(m,1)y}$	$a^{(m,2)x}/a^{(m,2)y}$	$a^{(m,3)x}/a^{(m,3)y}$	...	$a^{(m,n)x}/a^{(m,n)y}$

Table 1. Example of data farm organization

Similarly to the farm data table, the BM and the CH genotyping pool analysis data were organized in a tabular way (see Table 2 and Table 3). The main difference in these tables with respect to Table 1 is that the information for the n-th STR is a vector including all the allele values obtained from the pool genotyping process.

Pool	STR1	STR2	STR3	...	STRn
BM	BMap1{...}	BMap2{...}	BMap3{...}	...	BMapn{...}

Table 2. Bulk Milk (BM) Example

Pool	STR1	STR2	STR3	...	STRn
CH	CHap1{...}	CHap2{...}	CHap3{...}	...	CHapn{...}

Table 3. Cheese (CH) Example

Additionally, the absolute RFU alleles peak for the full farm (cows), BM and CH was organized according to Table 4.

Due to the complexity of processing raw data, in this preliminary study data from 3 out of the 12 months of sampling have been analyzed. At the end all tabular data were stored in plain text files using comma-separated values (CSV) format.

	STR1	STR2	STR3	...	STRn
COW1	$h^{(1,1)x}/h^{(1,1)y}$	$h^{(1,2)x}/h^{(1,2)y}$	$h^{(1,3)x}/h^{(1,3)y}$	...	$h^{(1,n)x}/h^{(1,n)y}$
COW2	$h^{(2,1)x}/h^{(2,1)y}$	$h^{(2,2)x}/h^{(2,2)y}$	$h^{(2,3)x}/h^{(2,3)y}$	...	$h^{(2,n)x}/h^{(2,n)y}$
COW3	$h^{(3,1)x}/h^{(3,1)y}$	$h^{(3,2)x}/h^{(3,2)y}$	$h^{(3,3)x}/h^{(3,3)y}$	...	$h^{(3,n)x}/h^{(3,n)y}$
...	...	...	...	...	...
COWm	$h^{(m,1)x}/h^{(m,1)y}$	$h^{(m,2)x}/h^{(m,2)y}$	$h^{(m,3)x}/h^{(m,3)y}$	...	$h^{(m,n)x}/h^{(m,n)y}$
BMh	BMhp1{...}	BMhp2{...}	BMhp3{...}	...	BMhpn{...}
CHh	CHhp1{...}	CHhp2{...}	CHhp3{...}	...	CHhpn{...}

Table 4. This table follows the same criteria of Tables 1, 2, and3 with the only difference that here we report the *height* of the RFU allele's peak.

2.2 Computer-Assisted Molecular Traceability

Background Motivation At first some sample data were tested by measuring the ability to trace the dairy products using well known software algorithms commonly used in genetic distance analysis like FSTAT [11], PHYLIP [10] and SMOGD [5] and then resorting to STRUCTURE [9] Bayesian algorithm approach.

However, functionalities offered by the aforementioned tools demonstrated to be not suited to accomplish the intended purpose. Therefore, we decided to implement a new approach able to detect if the BM or CH the's fingerprint could be traced and compared with the genetic pool characteristics of the pertinent farm.

The next subsection provides the reader with the general principles about the covariance matrix adaptation evolution strategy (CMA-ES) which is necessary to better understand the proposed computer-assisted molecular traceability method described next.

CMA-ES The covariance matrix adaptation evolution strategy (CMA-ES) is an optimization method first proposed by Hansen, Ostermeier, and Gawelczyk [8] in mid 90s, and further developed in subsequent years [7], [6].

Similar to quasi-Newton methods, the CMA-ES is a second-order approach estimating a positive definite matrix within an iterative procedure. More precisely, it exploits a *covariance matrix*, closely related to the inverse Hessian on convex-quadratic functions. The approach is best suited for difficult non-linear, non-convex, and non-separable problems, of at least moderate dimensionality (i.e., $n \in [10, 100]$).

In CMA-ES, iteration steps are called *generations* due to its biological foundations. The value of a generic algorithm parameter y during generation g is denoted with $y^{(g)}$. The mean vector $\mathbf{m}^{(g)} \in \mathbb{R}^n$ represents the favorite, most-promising solution so far. The *step size* $\sigma^{(g)} \in \mathbb{R}_+$ controls the step length, and the *covariance matrix* $\mathbf{C}^{(g)} \in \mathbb{R}^{n \times n}$ determines the shape of the distribution ellipsoid in the search space. Its goal is, loosely speaking, to fit the search

distribution to the contour lines of the objective function f to be minimized: $\mathbf{C}^{(0)} = \mathbf{I}$.

Most noticeably, the CMA-ES requires almost no parameter tuning for its application. The choice of strategy internal parameters is not left to the user. Notably, the default population size λ is comparatively small to allow for fast convergence. Restarts with increasing population size have been demonstrated [2] to be useful for improving the global search performance, and it is nowadays included as an option in the standard algorithm.

In this study we used the CMA-ES R package [17].

Implemented Procedure The proposed method is composed of two independent steps, the first one is the check-Test (cT) and the second one is the Genetic Pool Simulation (GPsim).

The goal of the cT is to verify, independently for each STR, the presence of all BMap$n$-th{...} or CHap$n$-th{...} elements from Table 2 and Table 3 in the STRn-th cow's farm list available in Table 1. The cT generates a score that assigns an increasing value proportional to the number of missed matches between BM or CH in the genetic pool. The so called penalty score (P) is given by the following formula:

$$P = \frac{n}{n_m} \tag{1}$$

where:

- n: is the total number of STRs available in the analysis
- n_m: is the STRs number in every BMap$n$-th{...} or CHap$n$-th{...} row with at least one allele missing in all the $\mathrm{a}^{(m\text{-th},n\text{-th})}$ cows alleles list.

P ranges from 0 to 1. The more P is close to zero, the more the following GPsim procedure should be considered reliable.

The GPsim procedure does not take into account the previously calculated P and assumes to have an estimate of the alleles amount (i.e., RFU peak in Table 4). GPsim assumes that, if a certain number of cows that produced the BM or CH do exist, then ideally it should exist a linear combination of alleles' cows able to recreate the genetic pool of BM and/or CH.

Starting from data organized as in Table 4, to search for linear combinations of the available data, normalization of each sample must be performed. Each sample is normalized between [0,1] by looking at the range of values available in the sample. As a result, a normalized set of data organized as in Table 5 is produced. Based on this normalized data set, an individual weight is computed for each cow to best fit the BMh or CHh fingerprint.

The CMA-ES is used to find the best weigh configuration in order to produce a weighted linear combination of cows as similar as possible to the genetic pool we want to trace. GPsim returns the mean square error (MSE) between the

Chn	STR1	STR2	STR3	...	STRn
COW1	$\frac{h^{(1,1)\mathrm{x}}}{max(h_1)} / \frac{h^{(1,1)\mathrm{y}}}{max(h_1)}$	$\frac{h^{(1,2)\mathrm{x}}}{max(h_1)} / \frac{h^{(1,2)\mathrm{y}}}{max(h_1)}$	$\frac{h^{(1,3)\mathrm{x}}}{max(h_1)} / \frac{h^{(1,3)\mathrm{y}}}{max(h_1)}$	...	$\frac{h^{(1,5)\mathrm{x}}}{max(h_1)} / \frac{h^{(1,5)\mathrm{y}}}{max(h_1)}$
COW2	$\frac{h^{(2,1)\mathrm{x}}}{max(h_2)} / \frac{h^{(2,1)\mathrm{y}}}{max(h_2)}$	$\frac{h^{(2,2)\mathrm{x}}}{max(h_2)} / \frac{h^{(2,2)\mathrm{y}}}{max(h_2)}$	$\frac{h^{(2,3)\mathrm{x}}}{max(h_2)} / \frac{h^{(2,3)\mathrm{y}}}{max(h_2)}$	...	$\frac{h^{(2,5)\mathrm{x}}}{max(h_2)} / \frac{h^{(2,5)\mathrm{y}}}{max(h_2)}$
COW3	$\frac{h^{(3,1)\mathrm{x}}}{max(h_3)} / \frac{h^{(3,1)\mathrm{y}}}{max(h_3)}$	$\frac{h^{(3,2)\mathrm{x}}}{max(h_3)} / \frac{h^{(3,2)\mathrm{y}}}{max(h_3)}$	$\frac{h^{(3,3)\mathrm{x}}}{max(h_3)} / \frac{h^{(3,3)\mathrm{y}}}{max(h_3)}$	...	$\frac{h^{(3,5)\mathrm{x}}}{max(h_3)} / \frac{h^{(3,5)\mathrm{y}}}{max(h_3)}$
...	...	...	...	...	...
COWm	$\frac{h^{(\mathrm{m},1)\mathrm{x}}}{max(h_m)} / \frac{h^{(\mathrm{m},1)\mathrm{y}}}{max(h_m)}$	$\frac{h^{(\mathrm{m},2)\mathrm{x}}}{max(h_m)} / \frac{h^{(\mathrm{m},2)\mathrm{y}}}{max(h_m)}$	$\frac{h^{(\mathrm{m},3)\mathrm{x}}}{max(h_m)} / \frac{h^{(\mathrm{m},3)\mathrm{y}}}{max(h_m)}$	...	$\frac{h^{(\mathrm{m},5)\mathrm{x}}}{max(h_m)} / \frac{h^{(\mathrm{m},5)\mathrm{y}}}{max(h_m)}$

Table 5. Normalized cows STR-RFU peak tabular data. The $max(h_{m-th})$ is the higher RFU peak for each cow.

expected milk or cheese alleles fingerprint and the predicted one, as described in the main GPsim steps below:

- Read the Chn and the BMh / CHh files → ***Chn***, ***BMh|CHh***
- Initialize the cows weight array equal to 0 → $\boldsymbol{W} = 0$
- Compute the cows weight array using the *CMA-ES*: cmaes(W, FitFunc, Chn, BMh|CHh,) → ***W****
- Normalize W* → ***Wn****
- Calculate the predicted BM|CH → **pBMh|pCHh** = Chn Wn*
- Normalize BMh|CHh and pBMh|pCHh → **BMhn|CHhn** and **pBMhn|pCHhn**
- Return the mean square error between BMhn-pBMhn or between CHhn-pCHhn ⇒ $\boldsymbol{MSE_{BM}|MSE_{CH}}$

The *CMA-ES* requires the definition of a fitness function (FitFunc) to be minimized. Essentially we try to minimize the difference between the BM—CH genetic profile and the corresponding profile computed as a linear combination of the cows' profiles:

FitFunc : FitFunc (tW, Chn, BMh|CHh) → ***fitness***

- Get the temporary cows weight array → **tW**
- Read the Chn, BMh|CHh files → ***Chn***, ***BMh|CHh***
- Calculate the temporary predicted BM|CH → **tBMh|tCHh** = Chn tW
- Return the fitness value ⇒ ***fitness*** = abs(BMh-tBMh)|abs(CHh-tCHh)

As a result of the full analysis we are able to use the available data providing two different outputs, P and MSE. When comparing diary products with cow profiles of the correct farm we expect to obtain P and MSE values close to 0. Increased values of P and MSE are instead indicators of presence of milk produced by cows not profiled in the considered farm and can be used as an indication of frauds.

3 Results

Figure 1 shows the CMA-ES genetic pool traceability simulation for Biella's data, while Figure 2 reports results for the Imperia's ones. Biella's dairy product

traceability has been tested over BM and CH for three different months: January, June and December 2013, while Imperia's dairy products traceability has been tested over BM and CH for February, July 2013 and February 2014.

The two figures show the ability of the proposed algorithm to reconstruct the BM, CH genetic fingerprint as a linear combination of the fingerprint of cows in the related farm, confirming the possibility of using the computed fingerprint as a traceability indicator for the products.

This is more evident in Figure 3 where an intentional cross-simulation has been performed. In this figure, one BM and one CH from Biella have been compared to cow's profiles from Imperia and viceversa. The figure shows that, the CMA-ES is unable to properly identify a linear combination of cow's profiles able to generate the analyzed BM and CH fingerprints, giving an indication of an anomaly.

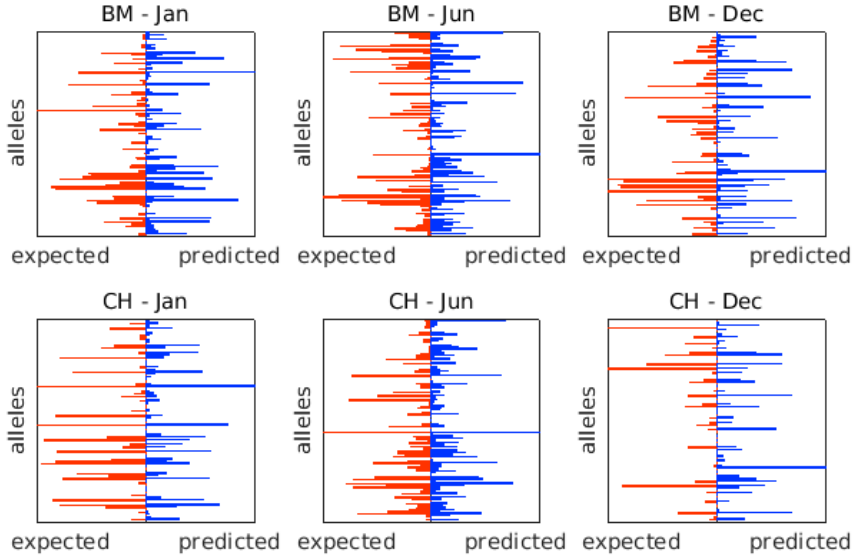


Fig. 1. Biella's traceability test. BM and CH are analyzed in three different months: January-June-December, 2013.

In order to summarize into a single numerical score the visual results proposed in Figures 1, 2 and 3 we propose a unique final score index (FS) that combines P and MSE together as follows:

$$FS = (1 - MSE)(1 - P) \tag{2}$$

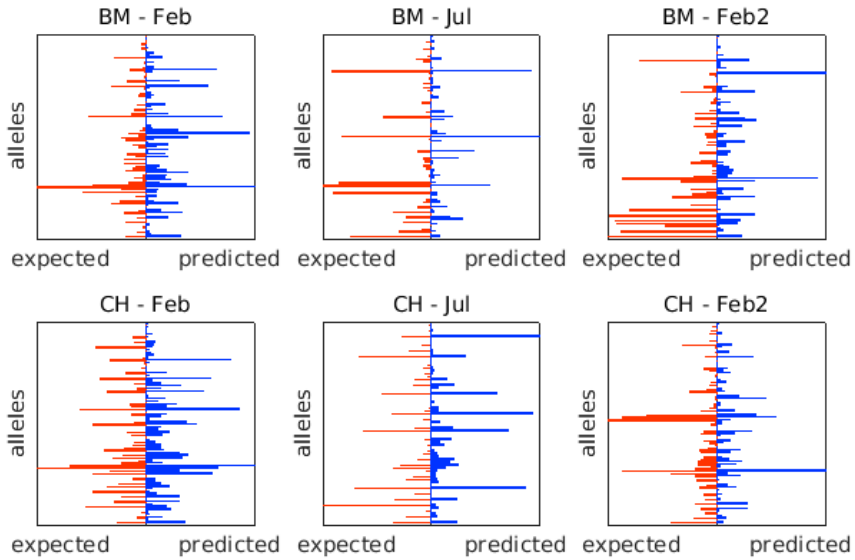


Fig. 2. Imperia's traceability test. BM and CH are analyzed in three different months: February-July, 2013 and February,2014.

Figure 4 summarizes all the aforementioned simulation results. The figure shows how FS can be used as a traceability indication for a diary product with respect to a given farm.

4 Conclusion

In this paper we proposed a computer-assisted molecular traceability system able to certify a traditional dairy product. We investigated the use of short tandem repeat analysis data processed by a Covariance Matrix Adaptation Evolution Strategy algorithm in order to predict the traceability between dairy products and the corresponding producer and to highlight possible adulterations and/or inconsistencies.

Preliminary results collected from two farms provided promising results. Additional data are still being processed in order to provide results on a wider and more complete dataset.

References

1. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: Basic local alignment search tool. Journal of Molecular Biology 215(3), 403 – 410 (1990), http://www.sciencedirect.com/science/article/pii/S0022283605803602

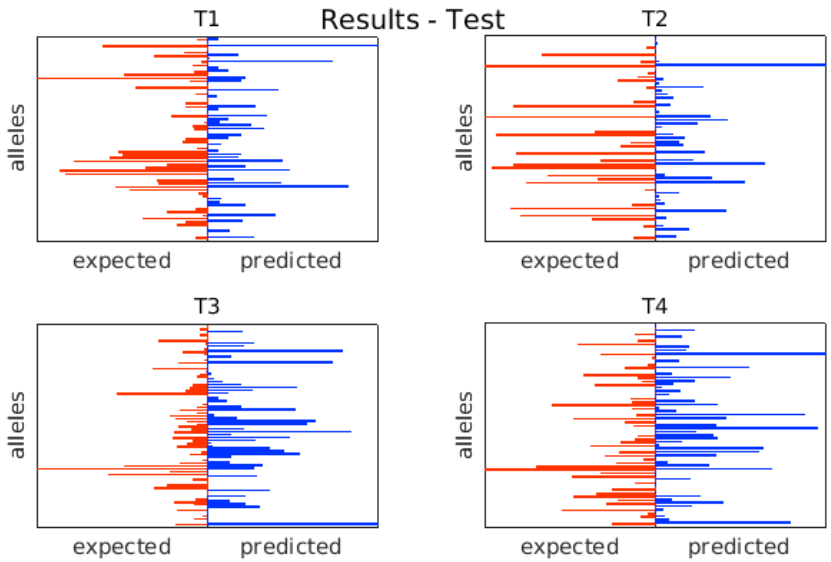


Fig. 3. Cross-simulation. T1: Biella's BM Jan,2013 with Imperia's farm. T2: Biella's CH Jan,2013 with Imperia's farms. T3: Imperia's BM Feb,2013 with Biella's farms. T4: Imperia's BM JFeb,2013 with Biella's farm.

2. Auger, A., Hansen, N.: A restart cma evolution strategy with increasing population size. In: Proc. IEEE Congress Evolutionary Computation. vol. 2, pp. 1769–1776 (2005)
3. Bagnato, A., Schiavini, F., Rossoni, A., Maltecca, C., Dolezal, M., Medugorac, I., Sölkner, J., Russo, V., Fontanesi, L., Friedmann, A., et al.: Quantitative trait loci affecting milk yield and protein percentage in a three-country brown swiss population. Journal of dairy science 91(2), 767–783 (2008)
4. Benson, D.A., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W.: Genbank. Nucleic Acids Research 43(D1), D30–D35 (2015), http://nar.oxfordjournals.org/content/43/D1/D30.abstract
5. Crawford, N.G.: smogd: software for the measurement of genetic diversity. Molecular Ecology Resources 10(3), 556–557 (2010), http://dx.doi.org/10.1111/j.1755-0998.2009.02801.x
6. Hansen, N., Müller, S.D., , Petrosnf, P.K.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). Evolutionary Computation 11, 1–18 (2003)
7. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation 9, 159–195 (2001)
8. Hansen, N., Ostermeier, A., Gawelczyk, A.: On the adaptation of arbitrary normal mutation distributions in evolution strategies: The generating set adaptation. In: Proceedings 6th International Conference on Genetic Algorithms. pp. 312–317. Morgan Kaufmann (1995)

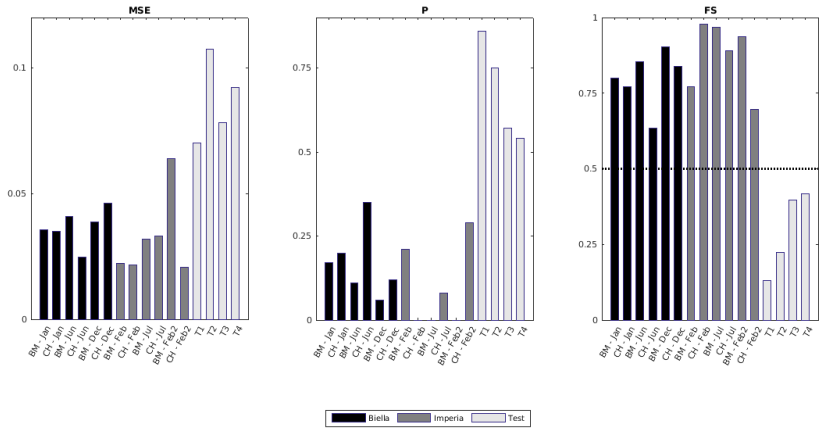


Fig. 4. Mean Square Error (MSE), Penalty (P) and Final Score (FS) results for traced dairy product and cross-simulation.

9. Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K.: Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9, 1322–1332 (2009)
10. J, F.: Phylip - phylogeny inference package (version 3.2). Cladistics 5, 164–166 (1989), http://evolution.genetics.washington.edu/phylip.html
11. J, G.: Fstat (v. 1.2): A computer program to calculate f-statistics. The Journal of Heredity 86(6), 485–486 (1995)
12. Lipkin, E., Mosig, M.O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A., Soller, M.: Quantitative trait locus mapping in dairy cattle by means of selective milk dna pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. Genetics 149(3), 1557–1567 (1998)
13. Megens, H.J., Crooijmans, R., San Cristobal, M., Hui, X., Li, N., Groenen, M.: Biodiversity of pig breeds from china and europe estimated from pooled dna samples: differences in microsatellite variation between two areas of domestication. Genetics Selection Evolution 40(1), 103–128 (2008)
14. Schnack, H.G., Bakker, S.C., van't Slot, R., Groot, B.M., Sinke, R.J., Kahn, R.S., Pearson, P.L.: Accurate determination of microsatellite allele frequencies in pooled dna samples. European journal of human genetics 12(11), 925–934 (2004)
15. Skalski, G.T., Couch, C.R., Garber, A.F., Weir, B.S., Sullivan, C.V.: Evaluation of dna pooling for the estimation of microsatellite allele frequencies: a case study using striped bass (morone saxatilis). Genetics 173(2), 863–875 (2006)
16. Team, R.C.: R: A language and environment for statistical computing [internet]. vienna, austria: R foundation for statistical computing; 2013. Document freely available on the internet at: http://www. r-project. org (2015)
17. Trautmann, H., Mersmann, O., Arnu, D.: cmaes: Covariance matrix adapting evolutionary strategy. R package, URL http://cran. r-project. org/package= cmaes (2011)