Breed of goat affects the prediction accuracy of milk coagulation properties using Fourier-transform infrared spectroscopy

Breed of goat affects the prediction accuracy of milk coagulation properties using Fourier-transform infrared spectroscopy

Journal:	*Journal of Dairy Science*
Manuscript ID	JDS.2020-19491.R2
Article Type:	Research
Date Submitted by the Author:	n/a
Complete List of Authors:	Stocco, Giorgia; Università degli Studi di Parma, Dipartimento di Scienze Medico-Veterinarie Dadousis, Christos; Università degli Studi di Parma, Dipartimento di Scienze Medico-Veterinarie Vacca, Giuseppe Massimo; Università degli Studi di Sassari, Dipartimento di Biologia Animale Pazzola, Michele; University of Sassari, Department of Animal Biology Paschino, Pietro; Università degli Studi di Sassari, Dipartimento di Medicina Veterinaria Dettori, Maria Luisa; Università degli Studi di Sassari, Dipartimento di Medicina Veterinaria Ferragina, Alessandro; Teagasc Food Research Centre Ashtown, Department of Food Quality and Sensory Science, Cipolat Gotet, Claudio; Universita degli Studi di Parma Dipartimento di Scienze Medico-Veterinarie
Key Words:	goat, Fourier-transform infrared spectroscopy, coagulation, curd-firming

INTERPRETIVE SUMMARY

Breed of goat affects the prediction accuracy of milk coagulation properties using Fourier-transform infrared spectroscopy. *By Stocco et al., page 000.* The aims of this study were to assess the feasibility in predicting goat milk coagulation traits via Fourier-transform infrared (FTIR) spectroscopy and to quantify the effect of four breeds on the predictions accuracy of these traits. Two validation procedures, Cross-Validation (CV) and Stratified CV (SCV) were adopted. Results from CV suggested the potential inclusion of the predicted coagulation traits in the routine acquisition of spectra from individual milk samples, as a useful alternative to instrumental testing. Conversely, the low prediction accuracies observed for the SCV suggested that, when using a multi-breed dataset, it is important to consider the differences among breeds.

14 PREDICTION OF GOAT MILK COAGULATION TRAITS

15 **Breed of goat affects the prediction accuracy of milk coagulation properties using Fourier-**
16 **transform infrared spectroscopy**

17

18 **Giorgia Stocco,[1] Christos Dadousis,[1] Giuseppe Massimo Vacca,[2*] Michele Pazzola,[2] Pietro**
19 **Paschino,[2] Maria Luisa Dettori,[2] Alessandro Ferragina,[3] and Claudio Cipolat-Gotet[1]**

20

21 [1]Department of Veterinary Science, University of Parma, 43126 Parma, Italy

22 [2]Department of Veterinary Medicine, University of Sassari, 07100 Sassari, Italy

23 [3]Department of Food Quality and Sensory Science, Teagasc Food Research Centre, D15 KN3K
24 Dublin, Ireland

25

26 *Corresponding author: gmvacca@uniss.it

ABSTRACT

The prediction of traditional goat milk coagulation properties (MCP) and curd firmness over time (CF_t) parameters via Fourier-transform infrared (FTIR) spectroscopy can be of significant economic interest to the dairy industry and can contribute to the breeding objectives for the genetic improvement of dairy goat breeds. Therefore, the aims of this study were to: i) explore the variability of milk FTIR spectra from four goat breeds (Camosciata delle Alpi, Murciano-Granadina, Maltese, and Sarda), and to assess the possible discriminant power of milk FTIR spectra among breeds, ii) assess the viability to predict coagulation traits by using milk FTIR spectra, and iii) quantify the effect of the breed on the prediction accuracy of MCP and CF_t parameters. In total, 611 individual goat milk samples were used. Analysis of variance of measured MCP and CF_t parameters was carried out using a mixed model including the farm and pendulum as random factors, and breed, parity and DIM as fixed factors. Milk spectra for each goat were collected over the spectral range from wavenumber 5,011 to 925 × cm^{-1}. Discriminant analysis of principal components (DAPC) was used to assess the ability of FTIR spectra to identify breed of origin. A Bayesian model was used to calibrate equations for each coagulation trait. The accuracy of the model and the prediction equation was assessed by Cross-Validation (CV; 80% training and 20% testing set) and Stratified CV (SCV; three breeds in the training set, one breed in the testing set) procedures. Prediction accuracy was assessed by using coefficient of determination of validation (R^2_{VAL}), the root mean square error of validation ($RMSE_{VAL}$), and the ratio performance deviation (RPD). Moreover, measured and FTIR predicted traits were compared in the SCV procedure, by assessing their least square means for the breed effect, Pearson's correlations, and variance heteroscedasticity. Results evidenced the feasibility of using FTIR spectra and multivariate analyses to correctly assign milk samples to their breeds of origin. The R^2_{VAL} values obtained with the CV procedure were moderate to high for the majority of coagulation traits, with $RMSE_{VAL}$ and RPD values increasing as the coagulation process progresses from rennet addition. Predictions accuracy obtained with the SCV were strongly influenced by the breed, presenting general low values restricting a practical application. In addition, the low Pearson's

correlation coefficients of Sarda breed for all the traits analyzed, and the heteroscedastic variances of Camosciata delle Alpi, Murciano-Granadina, and Maltese breeds, further indicated that it is fundamental to consider the differences existing among breeds for the prediction of milk coagulation traits.

Key words: goat, Fourier-transform infrared spectroscopy, coagulation, curd-firming

INTRODUCTION

World milk production has been risen up by 1.6% in 2018 and is expected to grow at 1.7% each year by 2028. Although goat milk, together with sheep and camel, account of only ~4% of the global milk production, the expected increase in the organic farming might advantages small ruminants over the cows (OECD-FAO, 2019). Indeed, the global dairy goat industry is rapidly expanding and its potential is promising especially for low- and medium-income countries, although new investments are needed to integrate markets, research, and production facilities (Miller and Lu, 2019). Goat dairy products around the world consist of yogurt, fermented milk, curd and cheese. Europe produces 34% of world goat cheese, although it counts 1.3% of the global goat population (FAOSTAT, 2018). Among a variety of dairy productions, cheese-making is a complex procedure, with many environmental and animal factors involved in the milk-to-cheese process. A brief indication of cheese-making is provided by the milk coagulation properties (**MCP**). Albeit the extensive research on MCP from cattle and sheep, a thorough knowledge on goat milk is restricted to recent scientific literature (Vacca et al., 2018a; Barłowska et al., 2020; Roy et al., 2020).

Milk coagulation properties can be measured in several methods (e.g., mechanical, vibrational, optical; Klandar et al., 2007), as well as different approaches are employed to model the coagulation process. For example, several studies have modeled the dynamics of milk curdling (e.g., prediction of storage modulus) and intensity of the process (e.g., acidification rate constant) as a function of time, using rheometers (Esteve et al., 2001; Gustavsson et al., 2014), while others exploited all the curd firmness values available from mechanical lactodynamographic instruments, to model the entire coagulation pattern, with the possibility to provide additional coagulation traits (i.e., speed of curd-firmness, syneresis rate) (Bittante et al., 2013; Cipolat-Gotet et al., 2018).

However, although MCP allow for a simultaneous evaluation of a considerable large number of milk samples in a daily routine, they are not suitable for studies intended at population level. Cost and logistics are the main restrictions for a wide-scale application. However, the moderate heritability of the traits (~0.15 - 0.27 in cattle) marks MCP as candidate traits for selection in the breeding

programs (Dadousis et al., 2016). A potential solution to overcome the aforementioned limits, and make available MCP data at population level, can be the prediction of MCP via Fourier-transform infrared (**FTIR**) spectroscopy in the range of near- and mid-infrared wavelengths. This technique is widely used in many laboratories for routine analysis of milk components (De Marchi et al., 2014; ICAR, 2020), and recently new focus has been given in its use in dairy cattle for several traits, from milk to animal health, and the environment (Tiplady et al., 2020).

In brief, the FTIR spectroscopy is based on using different waves of the infrared region of the electromagnetic spectrum to excite molecules in milk in relation to their rotational and vibrational structure (Karoui et al., 2010). Therefore, the spectrum reflects the quantities of the various chemical bonds within the milk sample. However, although FTIR spectroscopy is effective for predicting traits directly measurable in milk (e.g., fat and protein %, fatty acids), when predicting indirect traits, such as milk processing characteristics (e.g., MCP, cheese-making traits), and to traits related to the animal condition (e.g., methane emissions, lameness), it must be taken into account that the nature of the prediction is primarily influenced by the relationship of these traits with the milk chemical components (e.g., relationship of the rennet coagulation time with milk protein). Among the indirect milk traits, MCP are of significant economic interest to the dairy industry due to their association with cheese production. Thus, their prediction via FTIR (**MCP_{IR}**) has been widely studied in bovine milk (Dal Zotto et al., 2008; De Marchi et al., 2013; Bonfatti et al., 2016). Less studies have considered MCP_{IR} in buffalo (Manuelian et al., 2017) and sheep (Ferragina et al., 2017; Cellesi et al., 2019), but none is available in the literature for the caprine species.

For the prediction of indirect milk traits, various aspects creating variation and influencing final inferences should be considered. To name some of them, the periodicity of instrument calibration with known-concentrations-milk samples, the dataset size to develop the calibration equation, the pre-processing and standardization of the spectra, the chemometric procedures and the quality of the calibration set and validation strategies adopted (Karoui et al., 2010; Tiplady et al., 2019; Tiplady et al., 2020). As regards to this last aspect, the cross-validation (**CV**) is a common

112 statistical procedure used to evaluate the performance of prediction equations of both direct (Rutten
113 et al., 2009) and indirect milk measures referred to the milk characteristics (McParland et al., 2011)
114 and animal condition (Bittante and Cipolat-Gotet, 2018). However, some authors have showed that
115 this approach tends to overestimate the accuracy of predictions (Qin et al., 2016; Roberts et al., 2017;
116 Wang and Bovenhuis, 2019), principally because the validation set could be partly dependent on the
117 calibration set (e.g., spectra from milk samples collected from: i) same animals in different times, ii)
118 different animals from the same farm and iii) same farm in different seasons; used in both calibration
119 and validation set). Moreover, the use of CV cannot detect any presence of a specific subpopulation
120 in the data (e.g., spectra collected from different farming systems, breeds, seasons, FTIR instruments).
121 For their application at an industrial level, breed effect is one of the most important aspects
122 to consider while building calibrations from individual samples. It is acknowledged that breed is the
123 second most important genetic feature, after species of ruminants, influencing milk composition and
124 coagulation properties (Bittante et al., 2012; Stocco et al., 2017; Vacca et al., 2018a). As a
125 consequence, the degree of absorption bands related to the milk components varies among species
126 (Nicolaou et al., 2010) and among breeds within species (Zaalberg et al., 2019), resulting in different
127 milk spectra. Nevertheless, scientific support on the contribution of the breed on the prediction
128 accuracy of indirect milk traits by using FTIR spectra is still limited. Few studies investigated the
129 FTIR prediction accuracies across breeds and those were related to milk fatty acids (Soyeurt et al.,
130 2011; Maurice-Van Eijndhoven et al., 2012). However, those authors did not study the effect of the
131 breed on the prediction accuracy of the tested traits.
132 As regards to caprine species, no studies have investigated the MCP_{IR} variability, the
133 prediction of curd firmness over time ($\mathbf{CF_t}$) parameters via FTIR ($\mathbf{CF_{tIR}}$), and neither assessed the
134 prediction accuracies of these phenotypes at the breed level. Therefore, the aims of this study were
135 to: i) explore the variability of four goat breeds milk FTIR spectra, and to assess the potential
136 discrimination of breed on the basis of FTIR, ii) assess the predictive performance of goat milk FTIR

spectra on coagulation traits by using milk FTIR spectra, and iii) quantify the effect of the breed on the prediction accuracy of MCP_{IR} and CF_{tIR} parameters using individual goat milk samples.

MATERIALS AND METHODS

Animals, Milk Sampling, Composition and Coagulation Properties

The study involved 611 goats from four breeds, two cosmopolitan (Camosciata delle Alpi, which is the Italian Alpine Chamois, N = 204; Murciano-Granadina, N = 142) and two local from Italy (Maltese, N = 121; Sarda, N = 144). A detailed description of these breeds and their characteristics are reported in Vacca et al. (2018a). Goats were reared in 19 farms distributed over the whole island of Sardinia (Italy). Farms were selected among those officially registered in the flock books and recording system of provincial associations of goat breeders, with an average flock size of 32 ± 10 goats. Farms were characterized by three different management systems: traditional or extensive (with free grazing of natural pastures, seasonal milk production, family operated, N = 5), intermediate or semi-extensive (with cultivated grasslands, control of estrus and kidding season, N = 8), and modern or semi-intensive system (with modern buildings and facilities, common use of TMR, out-of-season kidding and continuous milk production, N = 6). Individual milk samples (200 mL/goat) were collected during the afternoon milking (one sampling day for each farm), stored at 4°C and analyzed within 24 h after collection. Daily milk yield was recorded as the total yield of morning plus evening milking of the same day of sampling.

The MilkoScan FT6000 (Foss, Hillerød, Denmark) was used to analyze the percentage of fat and protein for each individual milk sample (ISO-IDF), over the spectral range from wavenumber 5,011 to 925 × cm^{-1}. As water represents the major constituent of milk, and the transmittance spectrum of milk (T) is very similar to that of water (Kaylegian et al., 2009), water transmittance can cover T. Usually T is lower than that of water because of the presence of other components, thus T is <1. However, T can also be >1 for the wavelengths not much affected by other components because the quantity of water in milk is less than 100%. As the concentration of a given substance in milk is

proportional to the radiation absorbance (A), this is calculated from zeroed transmittance as A = log(1/T), when the transmittance of milk is equal to that of water, and thereby T = 1, and A = 0; when T <1, then A is positive; when T >1, then A is negative. The obtained absorbance spectra of milk samples corrected for water are automatically standardized by the instrument to correct the modifications in wavelength and/or absorbance scale. Two spectral acquisitions were carried out for each sample, and the results were averaged before data analysis. Somatic cell count (**SCC**) was measured by using a Fossomatic 5000 (Foss Electric A/S, Hillerod, Denmark), then log-transformed to somatic cell score [**SCS** = $\log_2(SCC \times 10^{-5}) + 3$; (Ali and Shook, 1980)]; total bacterial count was analyzed by a BactoScan FC150 analyzer (Foss Electric A/S, Hillerod, Denmark) and log-transformed [**LBC** = $\log_{10}$ (total bacterial count/1,000)].

Analysis of MCP (60 min test) was performed using the Formagraph instrument (Foss Electric A/S, Hillerod, Denmark). Rennet (Hansen Naturen Plus 215, Pacovis Amrein AG, Bern, Switzerland) was diluted in distilled water to obtain a solution at 1.2% (wt/vol), with final value of international milk clotting units (**IMCU**) of 0.0513 IMCU/milk mL. The recorded MCP were: rennet coagulation time (**RCT**, min), defined as the time interval between rennet addition and gelation; curd-firming time ($\mathbf{k_{20}}$, min), as the time between gelation and the attainment of curd firmness (**CF**) of 20 mm; CF at 30, 45 and 60 min after rennet addition ($\mathbf{a_{30}}$, $\mathbf{a_{45}}$, and $\mathbf{a_{60}}$, mm).

During testing, the instrument records the width (mm) of the oscillatory graph of the pendula every 15 seconds. Consequently, 240 individual values of CF were recorded for each milk sample in a 60 min analysis. The differences in each CF_t pattern of the individual samples were measured by the following 4-parameters model (Bittante et al., 2013):

$$CF_t = CF_P \times [1 - e^{-k_{CF} \times (t - RCTeq)}] \times e^{-k_{SR} \times (t - RCTeq)},$$

where CF_t is curd firmness at time t (mm); $\mathbf{CF_P}$ is the asymptotical potential value of CF at an infinite time in absence of syneresis (mm); $\mathbf{k_{CF}}$ is the curd-firming instant rate constant (%/min); $\mathbf{k_{SR}}$ is the syneresis instant rate constant (%/min); and $\mathbf{RCT_{eq}}$ is RCT estimated by CF_t equation on the basis of all data points (min). By using all the individual CF values, it is possible to derive other

two traits: the maximum CF ($\mathbf{CF_{max}}$, mm) achieved after a given time interval ($\mathbf{t_{max}}$, min). Values of the coagulation traits outside the interval of the mean ±3 standard deviations (**SD**) were excluded as outliers.

Milk yield, composition and coagulation traits (mean ± SD) for each breed have been provided as supplemental material (Supplemental Table S1).

FTIR Spectra and Statistical Analysis

Mixed Model

All the MCP and CF_t parameters were analyzed using a MIXED procedure (SAS Institute Inc., Cary, NC), according to the following model:

$$y_{mnopqr} = \mu + Farm_m + Breed_n + Parity_o + DIM_p + Pendulum_q + e_{mnopqr}$$

where y_{mnopqr} is the observed trait (RCT, k_{20}, a_{30}, a_{45}, and a_{60}; RCT_{eq}, k_{CF}, k_{SR}, CF_P, CF_{max}, and t_{max}); μ is the overall population mean; $Farm_m$ is the random effect of the m^{th} farm (m = 1 to 19); $Breed_n$ is the fixed effect of the n^{th} breed (n = Camosciata delle Alpi, Maltese, Murciano-Granadina, and Sarda); $Parity_o$ is the fixed effect of the o^{th} parity [o = 1 to 3; class 1: 1st and 2nd (209 samples); class 2: 3rd and 4th (194 samples); class 3: ≥5th (208 samples)]; DIM_p is the fixed effect of the p^{th} class of days in milk [p = 1 to 4; class 1: < 80 days (167 samples); class 2: 80-120 d (173 samples); class 3: 121-160 d (180 samples); class 4: >161 d (91 samples)]; $Pendulum_q$ is the random effect of the q^{th} measuring unit of the Formagraph instrument (q = 1 to 10); e_{mnopqr} is the random residual ~ N (0, σ_e^2).

Spectra Editing

Prior to spectra analysis, each single wavenumber of the spectra was standardized to a null mean and a unit sample variance. Mahalanobis distances were calculated by means of the Mahalanobis function of the R software, the inverse of the spectral covariance matrix and the "center" statement as a vector of 0. No outliers were recorded because all the spectra presented a distance value lower than the mean±3 standard deviations. Except the aforementioned standardization, the entire spectrum has been used and the spectra were not subjected to any mathematical pretreatment.

Breed traceability on the basis of FTIR spectra

Discriminant analysis of principal components (**DAPC**) was used to assess the ability of FTIR spectra (N = 611) to identify breed of origin by minimizing within breed variation, while optimizing the variance between breeds (Jombart et al., 2010). DAPC was performed using the adegenet R package (Jombart, 2008). Following this procedure, the 20 first PC were retained and they were used for the DAPC. The number of PC to be included in DAPC was inspected manually and a threshold of 95% of the original variability captured by PC was used.

Bayesian Model

Separate prediction models were fitted for all the MCP and CF_t traits. A Bayesian model (BayesB) was adopted as implemented in the BGLR R package (de los Campos and Perez Rodriguez, 2015). Details of this procedure are listed in Ferragina et al. (2015). Briefly, each MCP and CF_t trait was regressed on standardized spectra covariates using the following linear model:

$$y_i = \beta_0 + \sum_{j=1}^{1,060} x_{ij}\beta_j + \varepsilon_i ,$$

where y_i is the measured phenotype of the i^{th} sample, β_0 is an intercept, $\{x_{ij}\}$ are standardized FTIR wavelength data ($j = 1,...,1,060$), β_j are the effects of each of the wavelengths, and ε_i are model residuals assumed to be *iid* (independent and identically distributed) with normal distribution centered at zero with variance σ_ε^2. Given the above assumption, the conditional distribution of the data, given the effects and variance parameters, is:

$$P(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} N(\mu_i, \sigma_\varepsilon^2),$$

where $\boldsymbol{\theta}$ represents the collection of model parameters $\boldsymbol{\theta} = \{\beta_0, \boldsymbol{\beta}, \sigma_\varepsilon^2\}$, $N(\mu_i, \sigma_\varepsilon^2)$ is a normal distribution centered at $\mu_i = \beta_0 + \sum_{j=1}^{1,060} x_{ij}\beta_j$ and with variance σ_ε^2, and $\boldsymbol{\beta} = \{\beta_j\}$ is a vector containing the effects of the individual spectra-derived wavelengths. Specification of the Bayesian model is completed by assigning prior distribution to the unknowns, $\boldsymbol{\theta}$. The default values of the built-in BGLR rules were used for all the model's hyper-parameters, and the inferences were based on 30,000 iterations with a burn-in of 10,000.

Cross-Validation and Stratified Cross-Validation Procedures

For each trait, the accuracy of the model and the prediction equation were assessed by Cross-Validation (CV) and Stratified CV (**SCV**) procedures using the sample set with 611 single records.

In the CV procedure, data were split into a training set (80% of the total records), that was used to build the equation, and a testing set (20% of the total records), used as validation. The training-testing procedure was repeated 10 times for each trait, changing the training and testing set samples each time. The samples in the training and testing sets were randomly assigned, but, for each replicate, the testing set was composed by 25% of each of the four breeds.

In the SCV procedure, the training set was composed by records from 3 breeds, and the testing set was composed by the remaining breed.

Assessment of Prediction Accuracy

In the CV and SCV procedures, predictions accuracy was measured using coefficient of determination of validation (**R^2_{VAL}**) and the root mean square error of validation (**$RMSE_{VAL}$**). The ratio performance deviation (**RPD**), calculated as the ratio between SD and the $RMSE_{VAL}$, was used to compare our results with those from a previous study on sheep milk that used the same methodology (Ferragina et al., 2017), and to assess predictions accuracies among goat breeds in the SCV procedure. Coagulation traits with $R^2_{VAL} < 0.40$ in CV were not further presented as results in the SCV.

Moreover, in the SCV procedure, measured and FTIR predicted traits were compared assessing their: i) mean values (LSMeans for breed effect testing the aforementioned mixed model); ii) correlation (Pearson's correlations); iii) variance heteroscedasticity (Levene's test).

RESULTS AND DISCUSSION

Variability of Goat Milk Spectra

Descriptive statistics of milk yield, composition, traditional MCP and CF_t parameters are summarized in Table 1. These traits presented quite large variability (Coefficient of Variation from

15% to 86%, respectively for protein and k_{SR}), especially due to the different goat breeds sampled, as proved the results per each breed reported in Supplemental Table S1.

In Figure 1 is reported the number of clusters (4 breeds) of the population dataset from the milk FTIR spectra, with an average of concordance assignment of about 80%. The four breeds had a correct percentage of population assignment of 78.3, 76.2, 78.1 and 86.1%, respectively for Camosciata delle Alpi, Maltese, Murciano-Granadina and Sarda. The highest percentage of correct population assignment of Sarda goats was in part expected, as the diversity of this breed was previously evidenced not only for the greater fat and protein contents in milk (Vacca et al., 2018a), but also for the efficient milk coagulation, curd firming, syneresis (Pazzola et al., 2018) and overall cheese-making process, leading to lower fat and protein losses in the whey (Vacca et al., 2018b). Hence, the results from DAPC on FTIR spectra reflected the differences in milk composition among breeds (fat, protein and lactose), which were in part expected, since FTIR spectroscopy measures the vibrations of chemical bonds within functional groups, thus generating a spectrum. The characteristic absorption bands are associated with specific milk components. For example, C=O, C-N (the amide I; ~1,653 × cm^{-1}), N-H and C-N signals (amide II; ~1,567 × cm^{-1}) have been used for the estimation of protein; the C-O (triglyceride ester; ~1,175 × cm^{-1}), C=O group (~1,750 × cm^{-1}) and C-H (acyl chain; 3,000-2,800 × cm^{-1}) frequencies are commonly used to determine fat; the C-O and C-H stretch (1,100 and 1,000 × cm^{-1}) have been associated with lactose (Karoui et al., 2010).

Few studies investigated the feasibility of using FTIR spectra and multivariate analyses to correctly classify milk samples by their breeds of origin (Valenti et al., 2013; Salleh et al., 2019). In the study of Valenti et al. (2013), milk samples from three cattle breeds (Montbéliarde, Normande, and Holstein; in total 676 bulk milk samples) were analyzed using FTIR and NIR (near infrared) technology, obtaining a better discrimination of milk samples between Normande and Holstein breeds with FTIR spectra. Salleh et al. (2019) provided with clear discrimination among three goat breeds (Saanen, Jamnapari and Toggenburg), albeit at a limited sample size (N = 18 individual milk samples) to represent the variability of the population. The results obtained in our study suggest that the

discriminant analysis applied on FTIR spectra could be useful to differentiate milk of local breeds (e.g., Sarda) destined to dairy products from that of commercial breeds (e.g., Camosciata delle Alpi). It is generally acknowledged the high value of dairy products from local breeds usually associated with the superior milk quality (Damián et al., 2008; Paschino et al., 2020). Moreover, several other characteristics, such as human cultural heritage, environment, climate adaptation, vegetable and animal biodiversity result in an added value of the dairy products derived from local breeds (Sepe and Argüello, 2019). Hence, a fast, cheap and accurate method (such as FTIR) to detect breed or origin of milk samples might protect both producers and consumers from fraud. Indeed, Nicolaou et al. (2010) suggested the potential use of the FTIR spectroscopy and multivariate analysis for the detection of different milk species (bovine, caprine, and ovine) and quantification of the adulteration of caprine or ovine milk with bovine's in different mixtures.

Coagulation traits are strongly influenced by milk composition, being largely affected by milk fat and protein concentrations (Stocco et al., 2018), milk udder health indicators such as somatic cells and bacterial count (Leitner et al., 2016; Stocco et al., 2019). Among genetic factors (Damián et al., 2008; Devold et al., 2011), coagulation patterns greatly vary also among breeds (Stocco et al., 2017). The patterns of coagulation of the four goat breeds are reported in Figure 2. Clear differences were observed among breeds in terms of k_{20} and all CF traits, excluding the two similar patterns of Camosciata delle Alpi and Murciano-Granadina goats, in agreement with a previous study on the modeling of coagulation of these breeds (Pazzola et al., 2018). Curd-firming time varied, approximately, between 3 (Sarda) and 5 min (Camosciata delle Alpi). This trait represents the first step of the curd dehydration, by which milk components are recovered and concentrated in the cheese curd. In goat milk, k_{20} in a range between 2 and 4 min was defined as optimal for maximizing percentage cheese yield and the recovery of nutrients in the curd (Vacca et al., 2020). Thirty min after rennet addition, even larger differences were observed among breeds, with the lowest a_{30} values recorded for Maltese (35 mm) and the highest for Sarda breed (50 mm). The CF_{max} was achieved

317 almost at the same time (t_{max} between 32 and 41 min), but varied among breeds (from 38.7 mm for
318 Maltese, to 52.2 mm for Sarda breed).
319
320 ***Prediction Accuracy of Coagulation Traits in Goat Milk***
321 In this study, the entire spectrum was used to predict coagulation traits. As depicted in the
322 Supplemental Figure S1, five wavelength infrared regions can be observed: i) the transition area
323 between the short-wavelength infrared (**SWIR**) and ii) mid-wavelength infrared (**MWIR**) portions
324 of the electromagnetic spectrum (SWIR-MWIR region); iii) another very short region in the MWIR
325 part, named MWIR-2 region, and iv) the MWIR-1 (3,048 to 1,701 × cm^{-1}); finally, the mid-long
326 wavelength infrared (MWIR-**LWIR**; 1,582 to 930 × cm^{-1}) regions of the spectrum. It is a common
327 practice to remove spectral bands (e.g., water absorption area) prior to main analysis. However, as
328 previously demonstrated (Bittante and Cecchinato, 2013; Wang et al., 2016; Ferragina et al., 2017)
329 the spectral areas typical of the water absorption bands contain significant chemical and genetic
330 information, especially when individual samples are used. The inclusion of the water regions of the
331 spectrum in the prediction models could be considered as a limitation of this study. However, the
332 regression coefficients for the wavelengths in the water regions were always close to zero, in
333 agreement with the results found by Ferragina et al. (2017) for sheep milk.
334 As aforementioned, when FTIR spectra are used to predict indirect measures referred to milk
335 processing (e.g., MCP, cheese-making traits) or animal condition (e.g., pregnancy, lameness), the
336 accuracy of the prediction is strongly influenced by their correlation with milk components (e.g., fat,
337 protein and lactose). In fact, milk composition influences its coagulation properties, but modifications
338 in composition must be relevant for changes in milk coagulation traits, as FTIR spectroscopy is not
339 expected to detect differences in renneting patterns. However, the suitability of the predictions
340 depends not only on their accuracy, but also on their applications (e.g., breeding vs. monitoring
341 purposes). In Table 2 are reported the prediction statistics for the CV procedure for goat milk MCP_{IR}
342 and CF_{tIR} parameters. To compare data of the present study with the ovine species, results from

343 Ferragina et al. (2017) are also presented. The study by Ferragina et al. (2017) allows for a direct
344 comparison of our results, as i) ewes were sampled in the same environment (farms located in
345 Sardinia region), ii) MCP and CF_t parameters were measured in the same way (e.g., instrument, type
346 of rennet, IMCU), iii) similar prediction model was applied, and iv) the same CV procedure was
347 followed. The R^2_{VAL} for goat milk traits ranged from 0.42 to 0.68 within MCP_{IR}, and from 0.14 to
348 0.60 within CF_{tIR} parameters. It is worth noting that in goat the prediction accuracy increases as the
349 coagulation process progresses from rennet addition (R^2_{VAL} from 0.42 for RCT to 0.68 for a_{60}). This
350 trend probably derives from the particular features distinguishing goat milk coagulation respect to the
351 other species. Coagulation in goat milk is characterized by a long-lasting gel formation, and weak gel
352 structure forming after rennet addition (e.g., dispersion of coarse particles rather than a continuous
353 firming network), resulting in soft curd, that needs long time to strengthen and to entrap the other
354 milk constituents (Ould Eleya et al., 1995; Zhao et al., 2014; Roy et al., 2020). It is important to
355 consider that we used a standardized concentration of rennet for each goat milk sample, normally
356 used for bovine milk (Bittante et al., 2012), in order to make fair comparisons with several previous
357 studies. Because of the features characterizing goat coagulation process (e.g., long-lasting gel
358 formation, and weak gel structure), it is expected to have higher accurate measurement of CF traits,
359 that led to a high accuracy of their FTIR predictions (Caredda et al., 2016). The possibility to predict
360 coagulation traits in goat milk, especially those describing the second part of the coagulation pattern
361 (e.g., a_{45}, a_{60}, CF_{max}, CF_P), could be of particular interest for the goat dairy industry. As it has been
362 proposed, the slow speed in curd-firming, soft curd and low CF_P values directly impair percentage
363 cheese yield and the recovery of nutrients in the curd (Vacca et al., 2020).

364 Comparing these results with those from sheep (Table 2, study by Ferragina et al., 2017), the
365 R^2_{VAL} ranged from 0.28 to 0.69 within MCP_{IR}, and from 0.18 to 0.67 within CF_{tIR} parameters. In
366 sheep, the prediction accuracy tended to decrease as the coagulation progresses (R^2_{VAL} from 0.69 for
367 RCT to 0.28 for a_{60}), as opposed to goat. This derives from the low repeatability of the traits
368 describing the second part of the coagulation pattern. Moreover, the coagulation in sheep is much

faster than in goat, and often CF_{max} is achieved within 30 min. After reaching CF_{max}, the coagulation curve is characterized by high syneresis value (k_{SR} = 0.9 %/min; Ferragina et al., 2017). In general, the k_{SR} is a low-repeatable coagulation trait both in cattle and sheep (Stocco et al., 2017; Ferragina et al., 2017), describing the expulsion of the whey from the contracting coagulum (descending part of the CF_t pattern). During the lactodynamographic analysis, when the whey is expelled inside the small vat used for the test, the coagulum floats in the whey and the pendulum records a minor resistance. As a result, a lower curd firmness value is registered. However, it could be that the actual curd firmness continues to increase, even if the instrument records decreasing resistance.

In the case of bovine milk, a study by Ferragina et al. (2015) using the same FTIR prediction model and CV approach as in our study, showed R^2_{VAL} and $RMSE_{VAL}$ for RCT of 0.63 and 3.6, respectively. Although RCT was the only MPC analyzed in that study, it is generally recognized that bovine MCP are characterized by high variability (Cipolat-Gotet et al., 2012) and low repeatability in the second part of the CF_t pattern (from 57% for t_{max} to 71% for a_{60}), mainly because of the high incidence of late-coagulating (RCT > 30 min) milk samples (Stocco et al., 2017). This assumption is in agreement with the decreasing R^2_{VAL} and RPD values for MCP_{IR} reported in Holstein-Friesian cattle, with decreasing R^2_{VAL} from 0.76 (RCT) to 0.40 (a_{60}), and RPD from 2.03 to 1.26 (for RCT and a_{60}, respectively; De Marchi et al., 2013); in Italian Simmental cattle breed R^2_{VAL} of 0.69 and 0.21 and RPD of 1.81 to 1.14, for RCT and a_{60}, respectively (Bonfatti et al., 2016); and mixed-breed MCP (Holstein-Friesian, Jersey, Norwegian Red and crossbred), with R^2_{VAL} of 0.61 for RCT and 0.26 for a_{60}, and RPD from 1.59 to 1.16, respectively (Visentin et al., 2019). The same decreasing trend in the prediction accuracy of traditional MCP_{IR} was evidenced also in sheep by Cellesi et al. (2019), and in buffalo species by Manuelian et al. (2017) (lower R^2_{VAL} and higher $RMSE_{VAL}$ moving from RCT to a_{30} in both studies). It is important to mention that, although the methods for the MCP measurement and chemometrics procedures employed in those studies differ from our analysis, the results are still comparable, suggesting that goat MCP_{IR} and CF_{tIR} parameters could be used as a useful alternative to instrumental testing.

395

396 ***Prediction Accuracy of Milk Coagulation Traits Across Breeds***

397 In Table 3 are summarized the prediction statistics of MCP_{IR} and CF_{tIR} parameters across goat
398 breeds by using the SCV. Qin et al. (2016) stated that random CV underestimates the error of the
399 prediction equation when there are systematic differences between groups. However, although the
400 CV is a procedure routinely adopted, SCV would allow to evaluate model performance across breeds,
401 representing a more realistic picture of the model performance for a routine application, as it prevents
402 records from the same breed to end up in both the training and validation sets, and because its
403 performance is evaluated taking into account variation of coagulation traits across breeds. In the
404 present study, differences between groups of breeds were clearly evidenced, therefore the use of CV
405 could have led to misleading accuracies if we consider the variability among breeds observed in
406 Figure 1. Indeed, although the differences, in terms of RPD and $RMSE_{VAL}$ values, among Camosciata
407 delle Alpi, Maltese and Murciano-Granadina breeds were small, Sarda breed greatly differed,
408 showing the scantiest prediction accuracies (e.g., $RMSE_{VAL}$ from 1.2 to 28.7, for k_{20} and a_{60},
409 respectively). Comparing these results with those obtained with the CV (Table 2), it is clear that in
410 the SCV, the R^2_{VAL} was sharply decreased ($\leq$ 0.50 among breeds and traits). One explanation of this
411 decrease is the lower variability of the validation set (now made by only one breed) compared with
412 the calibration set (made by three breeds). In particular, in comparison with predictions derived from
413 CV, Sarda breed showed the greatest differences, with RPD value more than halved and $RMSE_{VAL}$
414 more than doubled in the case of a_{60}.

415 As previously mentioned, the negative effects of a slow curd-firming, weak gel structure, and
416 soft curd on cheese yield and recovery of nutrients in the curd (Vacca et al., 2020) suggests more
417 focus to be given to CF traits more than RCT, as the coagulation occurs longer after gelation in goat
418 milk. The possibility to implement MCP_{IR} and CF_{tIR} parameters rapidly and at individual level in the
419 routine milk recording system is of particular interest for the dairy goat cheese industry and breeding
420 associations. This could be particularly useful for those breeds with an important incidence of

intermediate (e.g., allele E; Alpine, Saanen, Toggenburg, Oberhasli, and LaMancha breeds), weak (e.g., allele F; Alpine, Saanen, Toggenburg, Oberhasli, LaMancha, and Nigerian Dwarf breeds), or non-expressing (e.g., allele N; Toggenburg and Nubian breeds) alleles of α_{s1}-casein, that have been associated with poor coagulation process (Maga et al., 2009; Devold et al., 2011).

Our results depicted the importance of the breed effect in MCP_{IR} and CF_{tIR}, that should be taken into account in future studies when predicting milk technological traits to avoid or correct misleading accuracies. The low prediction accuracies yielded for Sarda, compared to the rest of the breeds, is hypothesized to be a result of its different milk composition, and the very low variability of its coagulation traits (the smallest SD; Supplemental Table S1). These results are consistent with the DAPC findings, were Sarda had the highest correct breed assignment, as a results of different milk FTIR characteristics. Moreover, a narrow range in the variability of the reference values is known to negatively affect the predictability of the traits studied (Manley, 2014). Therefore, when building calibrations with samples from the three breeds validated on Sarda set, the result was the low accuracy of prediction for the latter. It is reasonable to state that the presence of Sarda goats in the calibration set could have affected the prediction accuracies for the other three breeds, even if to a lesser extent. A previous study attempting to predict milk fatty acids across four cattle breeds reported very high R^2_{VAL} for the majority of fatty acids examined (between 0.60 and 0.80; Maurice-Van Eijndhoven et al., 2012). Those authors developed calibrations from a multi-breed dataset (N = 1,236), validated on a multi breed external dataset (N = 190), without taking into account a single breed per validation procedure.

Effective phenotyping using FTIR based data is indeed dependent on the magnitude of the phenotypic correlations between the predicted vs. measured traits (Tiplady et al., 2020). The Pearson correlation coefficients for Sarda breed presented in Table 4 were low (< 0.4) for all traits analyzed; and although the correlation coefficients for Camosciata delle Alpi (from 0.49 to 0.69 for a_{60} and CF_P, respectively), Maltese (from 0.47 to 0.63 for a_{60} and k_{20} - CF_P, respectively), and Murciano-Granadina breeds (from 0.49 to 0.71 for a_{60} and CF_P, respectively) were higher, the heteroscedastic variances of

these three cosmopolite breeds assessed by Levene's test further indicated that the breed affected the prediction accuracy (Table 4). Again, this could be attributable to the lower variability of Sarda breed compared to the others. In fact, the narrow variability range of the traits from Sarda is comprised within the large variability range of the calibration set composed by the other three breeds, resulting in homoscedastic variance (non-significant Levene's test; excluding k_{20} and a_{60}). On the contrary, the inclusion of Sarda in the calibration set reduced the overall variability, leading to heteroscedastic variances in the other three breeds (except for a_{60} in Camosciata and Maltese breeds; Table 4).

Regarding the CF_{tFTIR} parameters, less differences among breeds in the CF_t patterns were observed (Figure 3), with the four CF_t curves converging closer compared to the measured CF_t (Figure 2). The narrow variability (after the introduction of the Sarda breed in the calibration set) led the small differences to become significant (data not shown), in terms of speed of curd-firming and syneresis rates. In fact, the shape of the curves in CF_{tIR} patterns of Camosciata and Murciano-Granadina breeds are steeper (described by k_{CF} trait) and more inclined (described by k_{SR} trait) (Figure 3), compared to the measured ones (Figure 2).

Opportunities and Possible Applications of Milk FTIR Spectra in Goats

Conventional goat breeding schemes are often hampered by the cost of measuring phenotypes and maintaining accurate data recording. Fourier-transform infrared spectroscopy has great potential for the future incorporation of traits that are hard or costly to measure in milk, as coagulation traits are, into breeding programs. However, except of prediction accuracy, the ability to successfully incorporate FTIR based coagulation properties into breeding programs is dependent on the heritability of the FTIR predicted traits, and on the genetic correlation between the predicted trait and the trait as measured by the standard reference method (Cecchinato et al., 2009). Those authors showed that MCP_{IR} could be used for genetic purposes even when the prediction accuracy values are moderate, as these traits are heritable and exhibit genetic correlations much higher than the phenotypic correlations with the corresponding measured traits. The results offered in the present

study are of absolute novelty for the goat dairy sector, and open new interest for improving our understanding of the genetics underlying the expression of FTIR predicted traits. Further, identification of genomic regions, for example via genome wide association studies, could provide with more insights on the biological basis of the traits (Gregersen et al., 2015; Dadousis et al., 2016; Dadousis et al., 2017). Establishing causal links between the genome and observed phenotypes may be assisted by employing the individual FTIR wavenumbers (Wang and Bovenhuis, 2018). Consolidating research towards these approaches would enable the future enhancement of goat dairy industry and breeders' associations.

CONCLUSIONS

Our results support the use of FTIR spectra to identify breed of origin of goat milk samples, a particularly important aspect for traditional dairy products and local breeds. Correct assignment is expected to increase by larger datasets than the one used in the present study.

Prediction accuracy values obtained with CV procedure were moderate to high for the majority of coagulation traits, suggesting their potential implementation in the routine acquisition of spectra from individual milk samples, as a useful alternative to instrumental testing. However, the negligible prediction accuracy based on SCV procedure confirmed that the accuracy of FTIR predictions was strongly influenced by how well the variation in the prediction population was represented in the calibration population. Ad hoc calibrations for the prediction of coagulation traits should be used for Sarda, as a result of different variability on milk components and coagulation compared to the rest of the breeds analyzed. When a multi-breed dataset is used, it is important to consider the differences existing among breeds. In this regard, different strategies in splitting the dataset for calibration and validation procedures should be tested that better reflect realistic scenarios applicable in the goat dairy industry. Further research should focus on the actual individual cheese yield, instead of the cheese-related MCP and CF_t.

ACKNOLEDGMENTS

Research supported by the Regional Government of Sardinia (Legge Regionale 7/2007; CUP J72I15000030007).

The authors thank the farmers for giving access to their flocks; the A.I.P.A./A.P.A.s (Provincial Farmers' Associations) of Cagliari, Nuoro, Sassari, and Oristano (Italy) and the firms Sepi Formaggi (Marrubiu, Italy) and L'Armentizia Moderna (Guspini, Italy) for their support in sample collection; and A.R.A. Sardegna (Regional Farmers' Association of Sardinia) for support in chemical milk analysis. The authors have not stated any conflicts of interest.

REFERENCES

Ali, A. and G. Shook. 1980. An optimum transformation for somatic cell concentration in milk. J. Dairy Sci. 63:487-490.

Barłowska, J., R. Pastuszka, J. Król, A. Brodziak, A. Teter, and Z. Litwińczuk. 2020. Differences in physico-chemical parameters of goat milk depending on breed type, physiological and environmental factors. Turk. J. Vet. Anim. Sci. 44.

Bittante, G., and A. Cecchinato. 2013. Genetic analysis of the Fourier transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. J. Dairy Sci. 96:5991-6006.

Bittante, G., and C. Cipolat-Gotet. 2018. Direct and indirect predictions of enteric methane daily production, yield, and intensity per unit of milk and cheese, from fatty acids and milk Fourier transform infrared spectra. J. Dairy Sci. 101:7219-7235.

Bittante, G., B. Contiero, and A. Cecchinato. 2013. Prolonged observation and modelling of milk coagulation, curd firming, and syneresis. Int. Dairy J. 29:115-123.

Bittante, G., M. Penasa, and A. Cecchinato. 2012. Invited review: Genetics and modeling of milk coagulation properties. J. Dairy Sci. 95:6843-6870.

524 Bonfatti, V., L. Degano, A. Menegoz, and P. Carnier. 2016. Short communication: Mid-infrared
525 spectroscopy prediction of fine milk composition and technological properties in Italian
526 Simmental. J. Dairy Sci. 99:8216-8221.
527 Caredda, M., M. Addis, I. Ibba, R. Leardi, M. F. Scintu, G. Piredda, and G. Sanna. 2016. Prediction
528 of fatty acid in sheep milk by midinfrared spectrometry with a selection of wavelengths by
529 genetic algorithms. LWT Food Sci. Technol. 65:503-510.
530 Cecchinato, A., M. De Marchi, L. Gallo, G. Bittante, and P. Carnier. 2009. Mid-infrared spectroscopy
531 predictions as indicator traits in breeding programs for enhanced coagulation properties of
532 milk. J. Dairy Sci. 92:5304-5313.
533 Cellesi, M., F. Correddu, M. G. Manca, J. Serdino, G. Gaspa, C. Dimauro, and N. P. P. Macciotta.
534 2019. Prediction of Milk Coagulation Properties and Individual Cheese Yield in Sheep Using
535 Partial Least Squares Regression. Animals. 9:663.
536 Cipolat-Gotet, C., A. Cecchinato, M. De Marchi, M. Penasa, and G. Bittante. 2012. Comparison
537 between mechanical and near-infrared methods for assessing coagulation properties of bovine
538 milk. J. Dairy Sci 95:6806-6819.
539 Cipolat-Gotet, C., M. Pazzola, A. Ferragina, A. Cecchinato, M. L. Dettori, and G. M. Vacca. 2018.
540 Technical Note: Improving modeling of coagulation, curd firming, and syneresis of sheep
541 milk. J. Dairy Sci. 101:5832-5837.
542 Dadousis, C., S. Biffani, C. Cipolat-Gotet, E. Nicolazzi, A. Rossoni, E. Santus, G. Bittante, and A.
543 Cecchinato. 2016. Genome-wide association of coagulation properties, curd firmness
544 modeling, protein percentage, and acidity in milk from Brown Swiss cows. J. Dairy Sci.
545 99:3654-3666.
546 Dadousis, C., S. Pegolo, G. J. Rosa, D. Gianola, G. Bittante, and A. Cecchinato. 2017. Pathway-based
547 genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield,
548 and curd nutrient recovery in dairy cattle. J. Dairy Sci.100:1223-1231.

549 Dal Zotto, R., M. De Marchi, A. Cecchinato, M. Penasa, M. Cassandro, P. Carnier, L. Gallo, and G.
550 Bittante. 2008. Reproducibility and repeatability of measures of milk coagulation properties
551 and predictive ability of mid-infrared reflectance spectroscopy. J. Dairy Sci. 91:4103-4112.
552 Damián, J. P., I. Sacchi, S. Reginensi, D. De Lima, and J. Bermúdez. 2008. Cheese yield, casein
553 fractions and major components of milk of Saanen and Anglo-Nubian dairy goats. Arq. Bras.
554 Med. Vet. Zootec. 60:1564-1569.
555 de los Campos, G., and P. Perez Rodriguez. 2015. BGLR: Bayesian Generalized Linear Regression.
556 R package version 1.0.4. Accessed June 10, 2019. http://CRAN.R-
557 project.org/package=BGLR.
558 De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2013. Prediction of coagulating and
559 noncoagulating milk samples using mid-infrared spectroscopy. J. Dairy Sci. 96:4707-4715.
560 De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared
561 spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171-1186.
562 Devold, T. G., R. Nordbø, T. Langsrud, C. Svenning, M. J. Brovold, E. S. Sørensen, B. Christensen,
563 T. Ådnøy, and G. E. Vegarud. 2011. Extreme frequencies of the αs1-casein "null" variant in
564 milk from Norwegian dairy goats - Implications for milk composition, micellar size and
565 renneting properties. Dairy Sci. Technol. 91:39-51.
566 Esteve, C. L. C., J. A. Lucey, and E. M. W. Pires. 2001. Mathematical modeling of the formation of
567 rennet-induced gels by plant coagulants and chymosin. J. Dairy Res. 68:499-510.
568 FAOSTAT (Food and Agriculture Organization of the United Nations Statistics Division). 2014.
569 Statistical Database of the Food and Agriculture Organization of the United Nations.
570 Accessed August 4, 2020. http://www.fao.org/faostat.
571 Ferragina, A., C. Cipolat-Gotet, A. Cecchinato, M. Pazzola, M. L. Dettori, G. M. Vacca, and G.
572 Bittante. 2017. Prediction and repeatability of milk coagulation properties and curd-firming
573 modeling parameters of ovine milk using Fourier-transform infrared spectroscopy and
574 Bayesian models. J. Dairy Sci. 100:3526-3538.

575 Ferragina, A., G. de los Campos, A. I. Vazquez, A. Cecchinato, and G. Bittante. 2015. Bayesian
576 regression models outperform partial least squares methods for predicting milk components
577 and technological properties using infrared spectral data. J. Dairy Sci. 98:8133-8151.
578 Gregersen, V. R., F. Gustavsson, M. Glantz, O. F. Christensen, H. Stålhammar, A. Andrén, H.
579 Lindmark-Månsson, N. A. Poulsen, L. B. Larsen, and M. Paulsson. 2015. Bovine
580 chromosomal regions affecting rheological traits in rennet-induced skim milk gels. J. Dairy
581 Sci. 98:1261-1272.
582 Gustavsson, F., M. Glantz, N. A. Poulsen, L. Wadsö, H. Stålhammar, A. Andrén, H. Lindmark
583 Månsson, L. B. Larsen, M. Paulsson, and W. F. Fikse. 2014. Genetic parameters for rennet-
584 and acid-induced coagulation properties in milk from Swedish Red dairy cows. J. Dairy Sci.
585 97:5219-5229.
586 ICAR (International Committee for Animal Recording). 2020. Guidelines: Section 12 - Milk
587 Analysis. Accessed April 24, 2020. https://www.icar.org/index.php/icar-recording-
588 guidelines/.
589 Jombart, T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers.
590 Bioinformatics. 24:1403-5.
591 Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a
592 new method for the analysis of genetically structured populations. BMC Genet**.** 11**:**94.
593 Karoui, R., G. Downey, and C. Blecker. 2010. Mid-infrared spectroscopy coupled with
594 chemometrics: A tool for the analysis of intact food systems and the exploration of their
595 molecular structure-quality relationships - A review. Chem. Rev. 110:6144-6168.
596 Kaylegian, K. E., J. M. Lynch, J. R. Fleming, and D. M. Barbano. 2009. Influence of fatty acid chain
597 length and unsaturation on midinfrared milk analysis. J. Dairy Sci. 92:2485-2501.
598 Klandar, A. H., A. Lagaude, and D. Chevalier-Lucia. 2007. Assessment of the rennet coagulation of
599 skim milk: A comparison of methods. Int. Dairy J. 17:1151-1160.

Leitner, G., Y. Lavon, Z. Matzrafi, O. Benun, D. Bezman, and U. Merin. 2016. Somatic cell counts, chemical composition and coagulation properties of goat and sheep bulk tank milk. Int. Dairy J. 58:9-13.

Maga, E. A., P. Daftari, D. Kültz, and M. C. T. Penedo. 2009. Prevalence of αs1-casein genotypes in American dairy goats. J. Anim. Sci. 87:3464-3469.

Manley, M. 2014. Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. Chem. Soc. Rev. 43:8200-8214.

Manuelian, C. L., G. Visentin, C. Boselli, G. Giangolini, M. Cassandro, and M. De Marchi. 2017. Short communication: Prediction of milk coagulation and acidity traits in Mediterranean buffalo milk using Fourier-transform mid-infrared spectroscopy. J. Dairy Sci. 100:7083-7087.

Maurice-Van Eijndhoven, M. H. T., H. Soyeurt, F.Dehareng, and M. P. L. Calus. 2013. Validation of fatty acid predictions in milk using mid-infrared spectrometry across cattle breeds. Animal. 7:348-354.

McParland, S., G. Banos, E. Wall, M. P. Coffey, H. Soyeurt, R. F. Veerkamp, and D. P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94:3651-3661.

Miller, B. A., and C. D. Lu. 2019. Current status of global dairy goat production: An overview. Asian-australas. J. Anim. Sci. 32:1219-1232.

Nicolaou, N., Y. Xu, and R. Goodacre. 2010. Fourier transform infra-red spectroscopy and multivariate analysis for the detection and quantification of different milk species. J. Dairy Sci. 93:5651-5660.

OECD (Organisation for Economic Co-operation and Development) FAO (Food and Agriculture Organization). 2019. OECD-FAO Agricultural Outlook 2019-2028. OECD Publishing, Paris/Food and Agriculture Organization of the United Nations, Rome. Accessed May, 8 2020. https://doi.org/10.1787/agr_outlook-2019-en.

625 Ould Eleya, M. M., S. Desobry Banon, and J. Hardy. 1995. A comparative study of pH and
626 temperature effects on the acidic coagulation of milk from cows, goats, and sheep. J. Dairy
627 Sci. 78:2675-2682.
628 Paschino, P., G. Stocco, M. L. Dettori, M. Pazzola, M. L. Marongiu, C. E. Pilo, C. Cipolat-Gotet, and
629 G. M. Vacca. 2020. Characterization of milk composition, coagulation properties and cheese-
630 making ability of goats reared in extensive farms. J. Dairy Sci. 103:5830-5843.
631 Pazzola, M., G. Stocco, M. L. Dettori, C. Cipolat-Gotet, G. Bittante, and G. M. Vacca. 2018.
632 Modeling of coagulation, curd firming, and syneresis of goat milk. J. Dairy Sci. 101:7027-
633 7039.
634 Qin, L. X., H. C. Huang, and C. B. Begg. 2016. Cautionary note on using cross-validation for
635 molecular classification. J. Clin. Oncol. 34:3931-3938.
636 Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J.
637 Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F.
638 Dormann. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or
639 phylogenetic structure. Ecography 40:913-929.
640 Roy, D., A. Ye, P. J. Moughan, and H. Singh. 2020. Gelation of milks of different species (dairy
641 cattle, goat, sheep, red deer, and water buffalo) using glucono-δ-lactone and pepsin. J. Dairy
642 Sci. 103:5844-5862.
643 Rutten, M. J. M., H. Bovenhuis, K. A. Hettinga, H. J. F. Van Vanlenberg, and J. A. M. van Arendonk.
644 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk
645 samples collected in winter and summer. J. Dairy Sci. 92:6202-6209.
646 Salleh, N. A., J. Selamat, G. Y. Meng, F. Abas, N. N. Jambari, and A. Khatib. 2019. Fourier transform
647 infrared spectroscopy and multivariate analysis of milk from different goat breeds. Int. J Food
648 Prop. 22:1673-1683.
649 Sepe, L., and A. Argüello. 2019. Recent advances in dairy goat products. Asian-australas. J. Anim.
650 Sci. 32:1306-1320.

651 Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. P. Berry, M. Coffey, and P.
652 Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds,
653 production systems, and countries. J. Dairy Sci. 94:1657-1667.
654 Stocco, G., C. Cipolat-Gotet, T. Bobbo, A. Cecchinato, and G. Bittante. 2017. Breed of cow and herd
655 productivity affect milk composition and modeling of coagulation, curd firming and syneresis.
656 J. Dairy Sci. 100:129-145.
657 Stocco, G., M. Pazzola, M. L. Dettori, P. Paschino, G. Bittante, and G. M. Vacca. 2018. Effect of
658 composition on coagulation, curd firming and syneresis of goat milk. J. Dairy Sci. 101:9693-
659 9702.
660 Stocco, G., M. Pazzola, M. L. Dettori, A. Summer, C. Cipolat-Gotet, and G. M. Vacca. 2019.
661 Variation in caprine milk composition and coagulation as affected by udder health indicators.
662 Int. Dairy J. 98:9-16.
663 Tiplady, K. M., T. J. Lopdell, M. D. Littlejohn, and D. J. Garrick. 2020. The evolving role of Fourier-
664 transform mid-infrared spectroscopy in genetic improvement of dairy cattle. J. Anim. Sci.
665 Biotechno. 11:1-13.
666 Tiplady, K. M., R. G. Sherlock, M. D. Littlejohn, J. E. Pryce, S. R. Davis, D. J. Garrick, R. J. Spelman,
667 and B. L. Harris. 2019. Strategies for noise reduction and standardization of milk mid-infrared
668 spectra from dairy cattle. J. Dairy Sci. 102:6357-6372.
669 Vacca, G. M., G. Stocco, M. L. Dettori, G. Bittante, and M. Pazzola. 2020. Goat cheese yield and
670 recovery of fat, protein, and total solids in curd are affected by milk coagulation properties. J.
671 Dairy Sci. 103:1352-1365.
672 Vacca, G. M., G. Stocco, M. L. Dettori, E. Pira, G. Bittante, and M. Pazzola. 2018a. Milk yield,
673 quality and coagulation properties of six breeds of goats: Environmental and individual
674 variability. J. Dairy Sci. 101:7236-7247.

675 Vacca, G. M., G. Stocco, M. L. Dettori, A. Summer, C. Cipolat-Gotet, G. Bittante, and M. Pazzola.
676 2018b. Cheese yield, cheese-making efficiency, and daily production of 6 breeds of goats. J.
677 Dairy Sci. 101:7817-7832.
678 Valenti, B., B. Martin, D. Andueza, C. Leroux, C. Labonne, F. La-halle, H. Larroque, P. Brunschwig,
679 C. Lecomte, M. Brochard, and A. Ferlay. 2013. Infrared spectroscopic methods for the
680 discrimination of cows' milk according to the feeding system, cow breed and altitude of the
681 dairy farm. Int. Dairy J. 32:26-32.
682 Visentin, G., A. McDermott, S. McParland, D. P. Berry, O. A. Kenny, A. Brodkorb, M. A. Fenelon,
683 and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared
684 spectroscopy analysis in dairy cows. J. Dairy Sci. 98:6620-6629.
685 Wang, Q., and H. Bovenhuis. 2018. Genome-wide association study for milk infrared wavenumbers.
686 J. Dairy Sci. 101:2260-2272.
687 Wang, Q., and H. Bovenhuis. 2019. Validation strategy can result in an overoptimistic view of the
688 ability of milk infrared spectra to predict methane emission of dairy cattle. J. Dairy Sci.
689 102:6288-6295.
690 Wang, Q., A. Hulzebosh, and H. Bovenhuis. 2016. Genetic and environmental variation in bovine
691 milk infrared spectra. J. Dairy Sci. 99:6793-6803.
692 Zaalberg, R. M., N. Shetty, L. Janss, and A. J. Buitenhuis. 2019. Genetic analysis of Fourier transform
693 infrared milk spectra in Danish Holstein and Danish Jersey. J. Dairy Sci. 102:503-510.
694 Zhao, L., S. Zhang, H. Uluko, L. Liu, J. Lu, H. Xue, F. Kong, and J. Lv. 2014. Effect of ultrasound
695 pretreatment on rennet-induced coagulation properties of goat's milk. Food Chem. 165:167-
696 174.

TABLES AND FIGURES

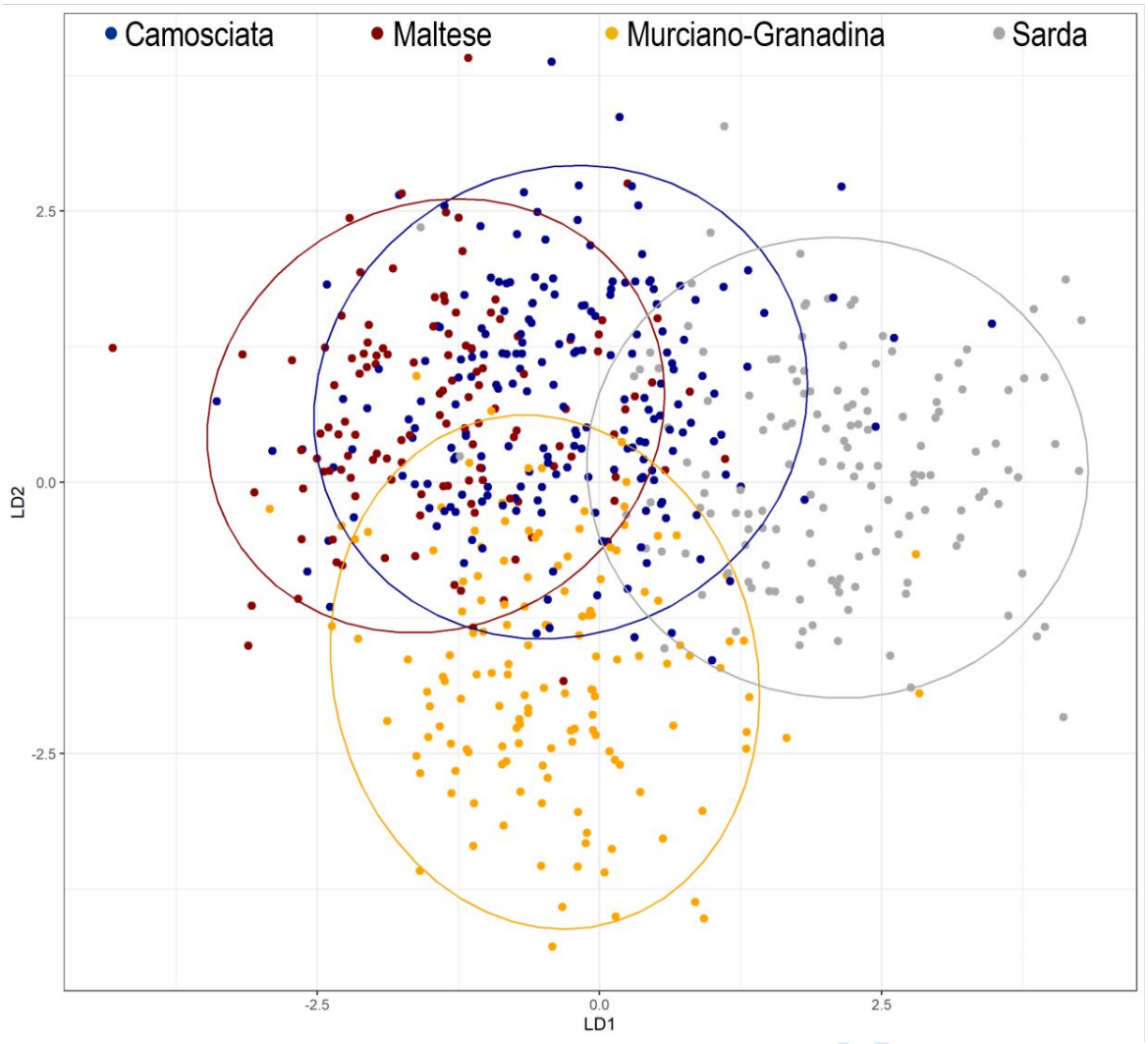
Table 1. Descriptive statistics of milk yield, composition, traditional milk coagulation properties (MCP) and curd firmness over time (CF_t) parameters of the 611 goat milk samples

Trait	Mean	SD	Min	Max	CoeffV[1], %
Milk Yield, kg/d	2.00	1.13	0.10	5.11	57
Milk composition					
Fat, %	4.39	1.31	1.93	8.38	30
Protein, %	3.55	0.53	2.36	5.25	15
SCS[2]	5.76	2.07	0.44	11.2	36
LBC[3]	1.71	0.82	0.30	4.23	49
Traditional MCP[4]					
RCT, min	12.4	4.31	4.00	29.3	35
k_{20}, min	4.02	1.87	1.45	14.5	47
a_{30}, mm	38.8	11.3	2.86	67.0	29
a_{45}, mm	39.3	11.9	8.30	66.7	30
a_{60}, mm	24.1	20.0	1.16	67.0	83
CF_t parameters[5]					
RCT_{eq}, min	13.1	4.27	4.24	30.0	33
k_{CF}, %/min	18.1	8.40	6.71	54.7	47
k_{SR}, %/min	0.58	0.50	0.15	2.91	86
CF_p, mm	47.8	11.2	13.7	75.3	23
CF_{max}, mm	42.3	9.91	12.1	66.7	23
t_{max}, min	38.3	11.8	12.8	60.0	31

[1]CoeffV = Coefficient of Variation; [2]SCS = $\log_2(SCC \times 10^{-5}) + 3$; [3]Logarithmic bacterial count (LBC) = $\log_{10}$(total bacterial count/1,000); [4]RCT = measured rennet gelation time; k_{20} = time interval between gelation and attainment of curd firmness of 20 mm; a_{30} (a_{45}, a_{60}) = curd firmness after 30 (45, 60) min from rennet addition; [5]RCT_{eq} = rennet coagulation time estimated by CF_t modeling; k_{CF} = curd firming instant rate constant; CF_P = asymptotic potential curd firmness; k_{SR} = syneresis instant rate constant; CF_{max} = maximum curd firmness achieved within 45 min; t_{max} = time at achievement of CF_{max}.

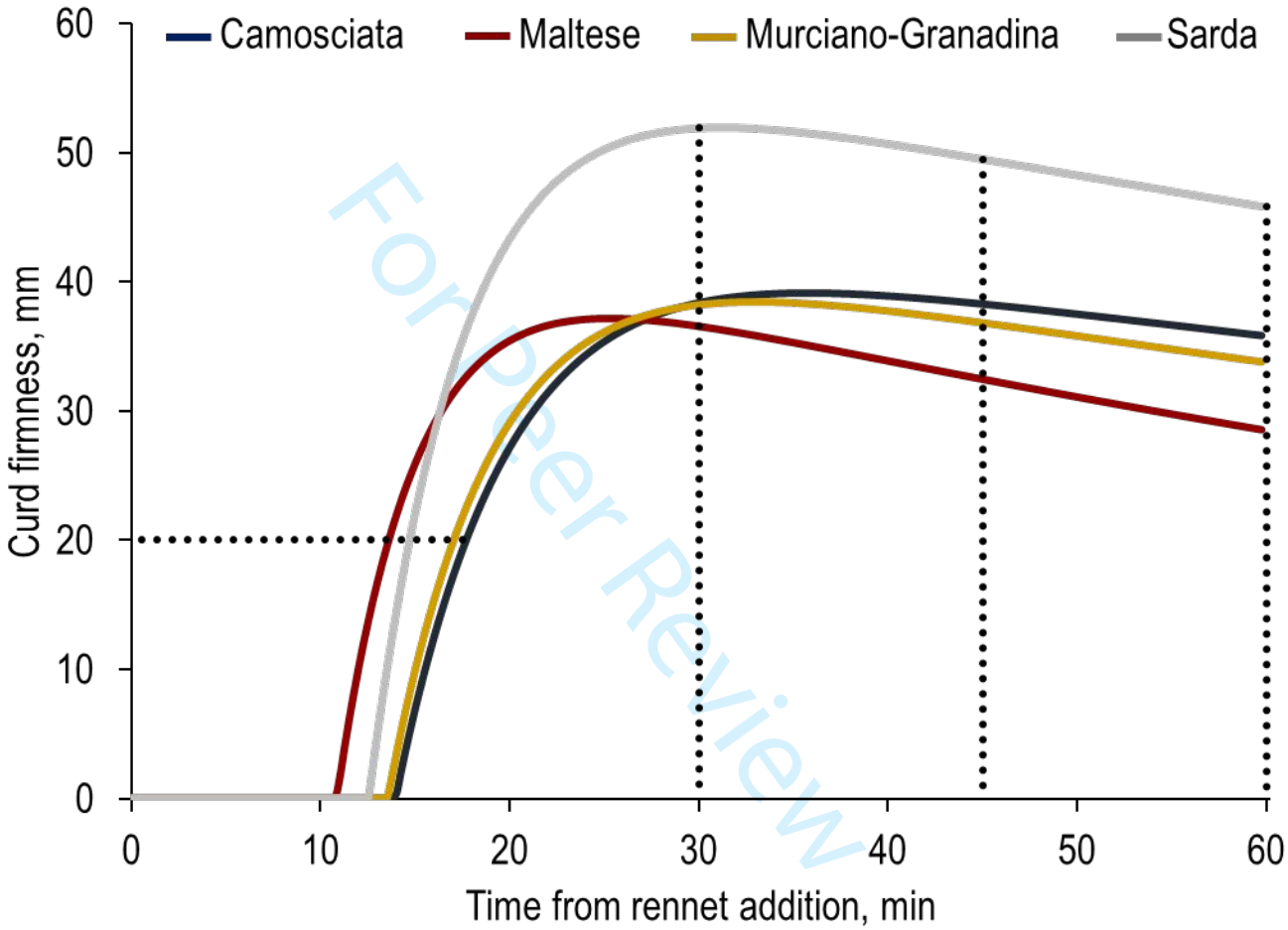
707 **Figure 1.** Discriminant analysis of principal components (DAPC) obtained from milk FTIR spectra

708 for the four goat breeds



709

710 **Figure 2.** Patterns of the measured curd firmness over time (CF_t) parameters of milk samples for the
711 four goat breeds. The intersection of the horizontal black dashed line and of the vertical black dashed
712 line at 30, 45 and 60 min with firmness curves represents k_{20} (the time from coagulation to a curd
713 firmness of 20 mm), a_{30} (curd firmness 30 min after rennet addition), a_{45} (curd firmness 45 min after
714 rennet addition) and a_{60} (curd firmness 60 min after rennet addition) of milk samples, respectively



715

Table 2. Prediction statistics[1] obtained with the Cross-Validation (CV) procedure for FTIR predicted traditional milk coagulation properties (MCP_{IR}) and curd firmness over time (CF_{tIR}) parameters obtained for goat milk in the present study, and for sheep milk in the study by Ferragina et al. (2017)

	Goat (N = 611) (present study)			Sheep (N = 1,089) (Ferragina et al., 2017)		
	R^2_{VAL}	$RMSE_{VAL}$	RPD	R^2_{VAL}	$RMSE_{VAL}$	RPD
Traditional MCP_{IR}[2]						
RCT, min	0.42	3.3	1.3	0.69	2.3	1.7
k_{20}, min	0.47	1.3	1.4	0.45	0.4	1.3
a_{30}, mm	0.48	8.4	1.4	0.48	9.0	1.2
a_{45}, mm	0.42	9.4	1.4	0.34	11.9	1.2
a_{60}, mm	0.68	11.5	1.8	0.28	13.8	1.2
CF_{tIR} parameters[3]						
RCT_{eq}, min	0.46	3.1	1.3	0.67	2.4	1.7
k_{CF}, %/min	0.15	8.0	1.1	0.23	10.4	1.1
k_{SR}, %/min	0.14	0.5	1.1	0.18	0.6	1.1
CF_p, mm	0.60	7.5	1.6	0.48	7.3	1.4
CF_{max}, mm	0.59	6.7	1.5	0.48	6.5	1.4
t_{max}, min	0.19	10.5	1.1	0.28	8.1	1.2

[1]R^2_{VAL} = coefficient of correlation of validation; $RMSE_{VAL}$ = root mean square error of validation; RPD = ratio performance deviation; [2]RCT = measured rennet gelation time; k_{20} = time interval between gelation and attainment of curd firmness of 20 mm; a_{30} (a_{45}, a_{60}) = curd firmness after 30 (45, 60) min from rennet addition; [3]RCT_{eq} = rennet coagulation time estimated by CF_t modeling; k_{CF} = curd firming instant rate constant; CF_P = asymptotic potential curd firmness; k_{SR} = syneresis instant rate constant; CF_{max} = maximum curd firmness achieved within 45 min; t_{max} = time at achievement of CF_{max}.

731 **Table 3.** Prediction statistics[1] across the four breeds obtained with the Stratified Cross-Validation (SCV) procedure for FTIR predicted traditional
732 milk coagulation properties (MCP_{IR}) and curd firmness over time (CF_{tIR}) parameters

	Camosciata delle Alpi			Maltese			Murciano-Granadina			Sarda		
	R^2_{VAL}	$RMSE_{VAL}$	RPD	R^2_{VAL}	$RMSE_{VAL}$	RPD	R^2_{VAL}	$RMSE_{VAL}$	RPD	R^2_{VAL}	$RMSE_{VAL}$	RPD
Traditional MCP_{IR}[2]												
RCT, min	0.27	4.1	1.1	0.25	3.8	1.1	0.32	4.1	1.1	0.13	3.7	0.9
k_{20}, min	0.34	1.9	1.1	0.39	1.3	1.3	0.34	1.4	1.2	0.03	1.2	0.7
a_{60}, mm	0.22	15.6	1.1	0.35	12.4	1.0	0.25	15.4	1.2	0.01	28.7	0.4
CF_{tIR} parameters[3]												
RCT_{eq}, min	0.29	4.0	1.1	0.28	3.6	1.2	0.38	4.1	1.1	0.10	4.9	0.7
CF_p, mm	0.47	7.8	1.2	0.39	8.0	1.3	0.50	7.2	1.4	0.33	8.0	1.0

733 [1]R^2_{VAL} = coefficient of correlation of validation; $RMSE_{VAL}$ = root mean square error of validation; RPD = ratio performance deviation; [2]RCT =
734 measured rennet gelation time; k_{20} = time interval between gelation and attainment of curd firmness of 20 mm; a_{60} = curd firmness after 60 min from
735 rennet addition; [3]RCT_{eq} = rennet coagulation time estimated by CF_t modeling; CF_P = asymptotic potential curd firmness.

736 **Table 4.** Pearson's correlations (r value and significance) and Levene's test between measured and
737 FTIR predicted coagulation traits

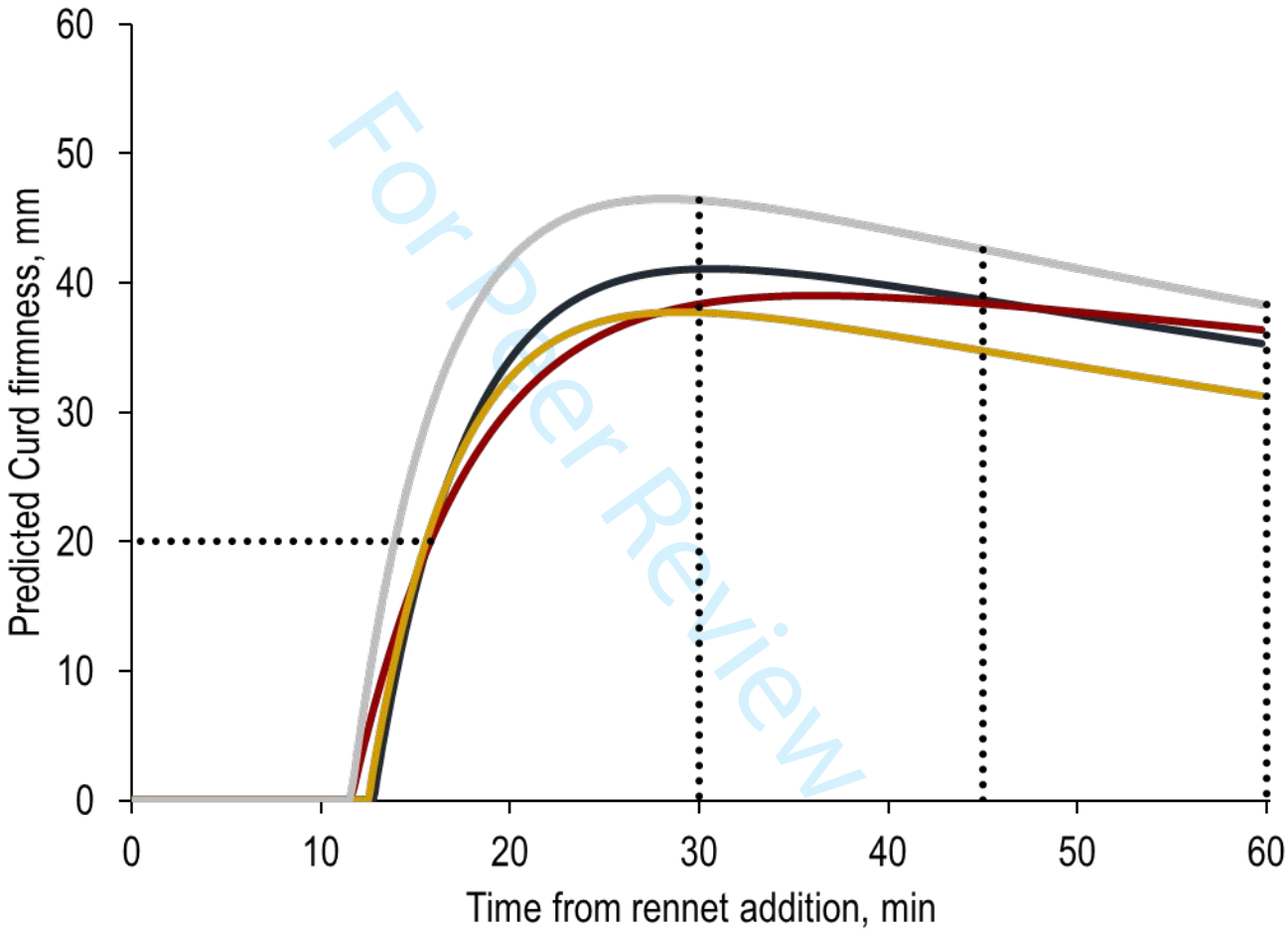
	Camosciata		Maltese		Murciano-Granadina		Sarda	
	r	Levene	r	Levene	r	Levene	r	Levene
Traditional MCP[1]								
RCT, min	0.53***	***	0.50***	**	0.57***	***	0.35***	
k_{20}, min	0.56***	***	0.63***	**	0.59***	*	0.33***	**
a_{60}, mm	0.49***		0.47***		0.49***	**	0.04	**
CF_t parameters[2]								
RCT_{eq}, min	0.54***	***	0.53***	**	0.62***	***	0.43***	
CF_P, mm	0.69***	***	0.63***	***	0.71***	**	0.57***	

738 [1]RCT = measured rennet gelation time; k_{20} = time interval between gelation and attainment of curd
739 firmness of 20 mm; a_{60} = curd firmness after 60 min from rennet addition; [2]RCT_{eq} = rennet
740 coagulation time estimated by CF_t modeling; CF_P = asymptotic potential curd firmness;
741 *** = $P < 0.001$; ** = $P < 0.01$; * = $P < 0.05$

742

743

744 **Figure 3.** Patterns of the FTIR predicted curd firmness over time (CF_{tIR}) parameters of milk samples
745 for the four goat breeds. The intersection of the horizontal black dashed line and of the vertical black
746 dashed line at 30, 45 and 60 min with firmness curves represents k_{20} (the time from coagulation to a
747 curd firmness of 20 mm), a_{30} (curd firmness 30 min after rennet addition), a_{45} (curd firmness 45 min
748 after rennet addition) and a_{60} (curd firmness 60 min after rennet addition) of milk samples,
749 respectively



750
751

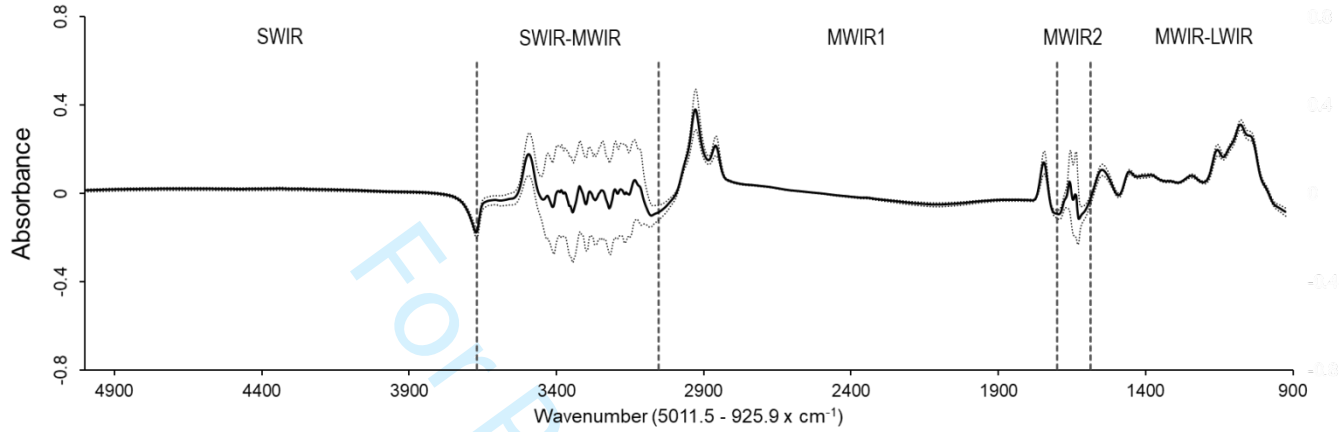
SUPPLEMENTAL MATERIAL

Supplemental Table S1. Descriptive statistics (mean±SD) of milk yield, composition, traditional milk coagulation properties (MCP), and curd firmness over time (CF_t) parameters for each breed of goat.

	Camosciata (n = 204)		Maltese (n = 121)		Murciano-Granadina (n = 142)		Sarda (n = 144)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Milk Yield, kg/d	2.80	1.01	1.54	1.02	2.33	0.85	0.94	0.33
Milk composition								
Fat, %	3.77	0.81	4.10	0.77	4.39	1.21	5.59	1.58
Protein, %	3.33	0.49	3.38	0.39	3.48	0.39	4.09	0.42
SCS[1]	5.22	2.22	6.53	1.93	5.56	1.92	6.07	1.89
LBC[2]	1.81	0.80	1.12	0.70	2.20	0.83	1.60	0.60
Traditional MCP[3]								
RCT, min	12.8	4.5	11.1	4.3	13.6	4.5	11.7	3.0
k_{20}, min	4.8	1.7	3.9	1.6	4.1	1.7	2.7	0.9
a_{30}, mm	35.7	7.8	35.6	10.1	37.4	9.5	49.5	8.0
a_{45}, mm	35.9	10.4	35	12.2	37.9	10.4	49.0	9.3
a_{60}, mm	17.6	17.1	10.4	12.5	22	17.8	46.1	11.1
CF_t parameters[3]								
RCT_{eq}, min	13.4	4.4	11.8	4.2	14.3	4.5	12.5	3.0
k_{CF}, %/min	14.7	5.4	21.0	11.3	17.4	7.1	19.6	6.8
k_{SR}, %/min	0.5	0.5	0.7	0.7	0.6	0.4	0.5	0.3
CF_p, mm	43.6	9.7	44.4	10.1	46.0	10.2	58.2	8.0
CF_{max}, mm	38.6	8.6	39.3	9.0	40.7	9.1	51.5	7.1
t_{max}, min	41.5	12.6	34.5	12.1	40.0	10.9	35.5	9.3

[1]SCS = $\log_2(SCC \times 10^{-5}) + 3$; [2]Logarithmic bacterial count (LBC) = $\log_{10}$(total bacterial count/1,000); [3]RCT = measured rennet gelation time; k_{20} = time interval between gelation and attainment of curd firmness of 20 mm; a_{30} (a_{45}, a_{60}) = curd firmness after 30 (45, 60) min from rennet addition; [4]RCT_{eq} = rennet coagulation time estimated by CF_t modeling; k_{CF} = curd firming instant rate constant; CF_P = asymptotic potential curd firmness; k_{SR} = syneresis instant rate constant; CF_{max} = maximum curd firmness achieved within 45 min; t_{max} = time at achievement of CF_{max}.

763 **Supplemental Figure S1.** Mean (solid line) ± SD (dotted line) of milk Fourier Transform infrared
764 (FTIR) goat milk spectra (range from 5,011 to 925 × cm^{-1}), and related spectral regions [short-
765 wavelength infrared (SWIR); short and mid-wavelength infrared (SWIR-MWIR); MWIR-1 and
766 MWIR-2; mid and long-wavelength infrared (MWIR-LWIR)].



767