Final report

Genomic selection against non-coagulating milk

Project number: O-15-20-274

Project period: 2015-07-01 -- 2018-06-30

Main applicant:
Assoc. Prof. Maria Glantz, Department of Food Technology, Engineering and Nutrition, Lund University
maria.glantz@food.lth.se

Co-applicants and project group members:
- PhD-student/MSc. Kajsa Nilsson, Department of Food Technology, Engineering and Nutrition, Lund University
- Post-doc/Dr. Sandrine Duchemin, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences
- Prof. Marie Paulsson, Department of Food Technology, Engineering and Nutrition, Lund University
- Prof. Dirk-Jan De Koning, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences
- Dr. Hans Stålhammar, VikingGenetics
- Dr. Freddy Fikse, Växa Sweden/Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences
- Dr. Mikka Stenholdt Hansen, Arla Foods amba
- Adj. Prof. Helena Lindmark Månsson, Dairy Sweden

Part 1: Detailed summary

Mjölkens förmåga att koagulera är avgörande för osttillverkning. Vi har tidigare visat att 31% av mjölken från SRB-kor har problem att koagulera: 13% är dåligt koagulerande och hela 18% är icke-koagulerande, vilket är alarmerande höga siffror. Syftet med detta projekt var därför att identifiera markörer och mutationer för icke-koagulerande mjölk för att genomiskt selektera mot denna oönskade egenskap i avelsprogram. Det övergripande målet är att minska frekvensen hos individer i aveln som ger upphov till icke-koagulerande mjölk och därmed avla kor med önskad genotyp som ger mjölk som resulterar i ost av hög kvalitet.

Mjölk- och blodprover har samlats in från 724 SRB-kor under stallperioden från 31 konventionella gårdar i Sverige. Proverna har analyserats för löpe-inducerad koagulering genom att använda två olika reologiska system, samt för mjölksammansättning och fysikaliska egenskaper genom användning av olika standardmetoder. Blodproverna användes för genetisk karakterisering, såsom finkartläggning av en potentiell genomisk region på BTA18, validering av

Projekt har fått finansiering genom:
Stiftelsen Lantbruksforskning

en genetisk markör (SNP) i VPS35-genen och för utvärdering av möjligheten för genomisk selektion mot icke-koagulerande mjölk. Datan har utvärderats med statistiska modeller med avseende på koaguleringsgrupp, besättning, laktationsnummer och laktationsstadie för de olika egenskaperna. Dessutom har kvantitativa statistiska analyser, med icke-koagulerande mjölk som en binär egenskap, utförts med sex olika statistiska modeller.

Projektet har bekräftat att 18.1% av mjölkproverna kan definieras som icke-koagulerande och ytterligare 18.9% av mjölkproverna som dåligt koagulerande. Detta resulterar i totalt 37% av mjölken från SRB-kor med försämrade koaguleringsegenskaper, vilket kan påverka industriell ostproduktion. Både icke-koagulererande och dåligt koagulerande mjölk fanns på alla de studerade gårdarna, vilket visar att den försämrade mjölken inte bara är ett problem på gårdsnivå utan snarare på nationell nivå. Vidare visade resultaten samband mellan icke-koagulerande mjölk och högre laktoshalt, mindre kasein i förhållande till totalt protein, lägre koncentrationer av både totalt kalcium och Ca^{2+} samt en högre titrerbarhet jämfört med koagulerande mjölk. Icke-koagulerande kan också korreleras med flera kemiska och fysikaliska egenskaper, men resultaten visade att ingen av de korrelerade parametrarna individuellt kan användas för att bedöma förekomsten av icke-koagulerande mjölk. Studier av koaguleringsprocessen visade att icke-koagulerande mjölk är kopplad till aggregeringssteget och inte det enzymatiska steget vid löpe-inducerad koagulering.

Finkartläggningen av en genomisk region på BTA18 visade att en QTL-region var starkt associerad med icke-koagulerande mjölk. Denna QTL förklarade 34% av den genetiska variationen för icke-koagulerande mjölk och baserades på 382 SRB-kor från 21 besättningar, vilka provtogs som en del i det Svensk-Danska Milk Genomics Initiativet (SLF-projekt V0930028/V1130006). En validering av denna QTL-region med genotypning av en SNP i VPS35-genen för alla kor utfördes genom att använda de 724 nya proverna från detta projekt. En effekt av QTLen i VPS35-genen kunde dock inte bekräftas i dessa nya prover. Det inkonsekventa resultatet för korna i det Svensk-Danska Milk Genomics Initiativet (SLF-projekt V0930028/V1130006) jämfört med kor från detta projekt kan indikera att resultaten från den tidigare studien av Duchemin *et al.* (2016) är ett falsk positivt resultat och att den genotypade SNPen i VPS35-genen kanske inte är en orsakande mutation för icke-koagulerande mjölk. En annan förklaring till att den tidigare funna OTLn inte kunde bekräftas kan vara att valideringsstudien hade för låg styrka. Vidare skattades starka genetiska korrelationer mellan icke-koagulerande mjölk och koaguleringsegenskaper i detta projekt. Dessa resultat tyder på att icke-koagulerande mjölk kan modifieras genom ett selektivt avelsarbete, utan att förändra mjölkens näringsmässiga egenskaper. Denna möjlighet undersöktes genom att analysera senarios för genomisk selektion, för vilka nogrannheten av skattningarna för GBLUP tyder på att genomisk selektion mot icke-koagulerande mjölk kan användas, om icke-koagulerande mjölk ingår i det nationella avelmålet för SRB-kor.

Resultaten från detta projekt gör det möjligt att utveckla en strategi för genomisk selektion mot icke-koagulerande mjölk, som kan användas för att begränsa användningen av tjurar som medför en ökad risk för icke-koagulerande mjölk. För ett direkt urval baserat på enskilda haplotyper eller gener krävs ytterligare valideringsstudier på ett större antal kor. Detta skulle kunna tillämpas i det nordiska avelsprogrammet och därigenom säkerställa en potentiell fördel för den svenska mejeri- och jordbruksindustrin samt andra intressenter. Genom att inkludera koaguleringsförmåga i avelsmålet kan egenskapen förbättras genom selektion och problemet med icke-koagulerande mjölk kan minskas. Dessutom kan en homogen mjölk med minimal variation i kvalitet levereras till de svenska mejerierna. Med en förbättrad selektion kan en ökad genetisk utveckling för mjölkkvalitetsegenskaper uppnås.

Part 2: The report (max 10 pages)

Introduction

The ability of milk to coagulate is crucial for cheese production. In previous studies, we have shown that 13% of milk from Swedish Red Dairy Cattle (RDC) is poorly coagulating (PC) and as much as 18% is non-coagulating (NC), which is an alarming high figure of 31% of the cows that give milk that has difficulties to coagulate (Gustavsson et al., 2014a, b; SLF-project V0930028; SLF-project V1130006). Furthermore, we have shown that genetic factors explain a large part of the phenotypic variation in milk coagulation ability (Gregersen et al., 2015; Gustavsson et al., 2014a,b) and also reported unfavorable genetic associations with milk yield and protein content (Gustavsson et al., 2014b). PC and NC milk will affect the industrial cheese production with lower cheese yield and quality as well as longer and more costly processing. Therefore, the coagulation ability of milk from Swedish RDC cows is of great importance for cheese production in Sweden, since the breed constitutes a large part of the national dairy cow population. If no actions are taken in the breeding program for the RDC breed, the problem of NC milk risks to increase.

The hypothesis is that NC milk can largely be explained by a limited number of genetic markers and that these markers are different for different cattle breeds. By using fine mapping to identify the regions of the cow genome that affect coagulation as well as the impact of milk composition, opportunities are given for genomic selection against the undesirable property. The aim of this project was therefore to identify markers and candidate mutations for NC milk from Swedish RDC in order to select against this undesirable property in the breeding program. For this purpose, data generated from the present project as well as from the Swedish-Danish Milk Genomics Initiative (SLF-projects V0930028/V1130006) has been used. The overall aim is to reduce the frequency of individuals in the breeding that give rise to NC milk and thus breed cows with the desired genotype that give milk that results in cheese of high quality. The present project has been divided into three workpackages (WP), including fine mapping of identified QTL-regions (quantitative trait loci) for non-coagulation (WP1), validation of QTL for non-coagulation (WP2) and effect of the genetic markers on milk composition (WP3).

Materials and methods

Milk and blood sample collection

Individual morning milk samples from 724 Swedish RDC were collected during the stable period between December 2015 to April 2016 and September 2016 to April 2017 from 31 conventional farms in the south of Sweden. The sampled cows were in parity 1-8 and between 10-436 days in milk (DIM), whereof 78% were in mid-lactation (70-280 DIM). All cows were milked twice a day and fed according to standard practices. The sampled cows represented the Swedish RDC population at the time of the study. Blood samples from the tail of the cows were also collected in connection to the milking.

In addition, data generated in the Swedish-Danish Milk Genomics Initiative (SLF-projects V0930028/V1130006) have been used together with data from the present project (SLF-project O-15-20-274) for part of the analyses.

Analyses of milk composition and physical traits

The milk samples have been analyzed for rheological, compositional and physical properties. Fat, protein, casein (CN), lactose and citric acid have been analyzed using Fourier Transform Infrared (FTIR), somatic cell count (SCC) using flow cytometry, total calcium using inductively coupled plasma atomic emission spectroscopy (ICP-AES), free Ca^{2+} using an ion selective electrode,

titratable acidity (TA) using titration with NaOH, casein micelle size using laser light scattering and genetic variants, phosphorylation and glycosylation of proteins using liquid chromatography-high resolution mass spectrometry (LC-HRMS). Additionally, capillary electrophoresis was used to determine where in the coagulation process the issue of NC milk is present by comparing coagulating milk samples and NC milk samples before and after rennet addition. All cows included in the milk analyses had a SCC < 300 000 cells mL^{-1}.

Analyses of milk coagulation

Rennet-induced coagulation was followed at 32°C and the skim milk samples were pH adjusted to 6.5 prior addition of chymosin to a final concentration of 0.09 IMCU mL^{-1} milk (Chy-Max Plus, 200 IMCU mL^{-1}, Christian Hansen A/S). The rheological properties were measured using two different rheological systems. The free oscillation rheometer (ReoRox4) was used to screen all milk samples, whereas a more detailed and sensitive rheological characterisation of randomly selected samples was measured using a low-amplitude rheometer (Stresstech) to verify the screening method. Parameters obtained from the two systems (rox and stress, respectively) were rennet coagulation time (RCT), gel strength (G'), curd firming rate (CFR) and yield stress (σ_y; resistance against gel breakdown). Milk samples were defined as NC and PC based on the mean RCT_{rox} of the two duplicates, see Table 1. For further quantitative analyses (fine-mapping of a QTL, validation of a QTL in the VPS35 gene, and genomic prediction) NC milk was scored as a binary trait: a) NC milk with the value 1 represented RCT >40 min, and b) NC milk with a value 0 represented RCT ≤ 40 min.

Isolation of DNA and the VPS35 genotypes

The extraction of DNA included 1,046 cows in total from the present project (N=646) and from the Swedish-Danish Milk Genomics Initiative (N=400; SLF-projects V0930028/V1130006). DNA from the blood samples were extracted at SLU using Qiasymphony SP® instrument (Qiagen, Hilden, Germany), for a total of 12 batches, and 96 samples per batch. With the extracted DNA, all cows were genotyped using StepOnePlus Real-Time PCR system (Life Technologies) together with a custom designed TaqMan SNP Genotyping Assays (Applied Biosystems). The genotype mutation is located on chromosome 18:15046826 in the VPS35 gene. From the 1,046 genotyped cows, 89 had no phenotypic records, 3 DNA samples were undetermined due to low amplification in the PCR, and these were excluded from the analyses. Finally, after genotyping, three data sets were created with different information for this same mutation: a) Real375, with 375 genotyped cows (generated from SLF-projects V0930028/V1130006); b) Imputed382, with 382 cows with imputed genotypes (generated from SLF-projects V0930028/V1130006); and c) Real582, with 582 genotyped cows for the VPS35 gene (generated from the present project). Genotype and allele frequencies of the SNP in the VPS35 gene are presented below in the Results and Discussion section.

Additional Genotypes

Different densities of genetic markers have been used in the project. These are: 1) a low-density SNP chip of 6,990 SNP (**LDSNP;** Infinium Bovine LD BeadChip from Illumina Inc., San Diego, CA), 2) a medium-density SNP chip of 50,000 SNP (**MDSNP;** Illumina BovineSNP50v1 BeadChip from Illumina), 3) a high-density SNP chip of 777,963 SNP (**HDSNP;** Illumina BovineHD BeadChip from Illumina), and 4) whole-genome sequences (**WGS**) of 429 sequenced individuals, including key ancestors of the Swedish RDC cow population available from the 3rd run of 1000 bulls genome consortium (Daetwyler et al., 2014). In addition, the composite genotypes of genetic variants of α_{s1}-, α_{s2}-, β-, and κ-caseins generated in the Swedish-Danish Milk Genomics Initiative (SLF-projects V0930028/V1130006) were used in this project.

Imputation

For the fine-mapping, three imputations were done from HDSNP to WGS, using different reference populations: a) 33 sequenced key-ancestors from Swedish RDC and Finnish Ayrshires breed, b) 284 sequenced animals belonging to 8 dairy breeds; and c) 429 sequenced animals belonging to all 15 breeds. With this approach, each variant was imputed three times based on the three different reference populations. The genotype with the highest imputation accuracy across the three imputations was selected as the best-imputed genotype, and combined in one data set.

For genomic prediction, the starting point was to use imputed genotypes provided by NAV (Nordic Cattle Genetic Evaluation). NAV's imputed MDSNP were obtained by imputing genotypes from LDSNP to MDSNP based on a large reference population of genotyped animals (more than 50,000). After this imputation, NAV imputed MDSNP was directly used and the MDSNP from the 392 cows genotyped with HDSNP (SLF-projects V0930028/V1130006) was extracted. The data set used for the genomic prediction study consisted of 1,046 Swedish RDC cows with imputed and genotyped MDSNP.

Statistical analyses

Regarding the statistical analysis of compositional and physical traits, mean values for each measurement were used to calculate mean values, standard errors and coefficients of variance for the measured milk traits. To evaluate means and influence from coagulation group, herd, parity and DIM on the different traits, the general linear model [1] was used:

$$Y_{ijk} = \mu + coag_i + herd_j + parity_k + b_1 \cdot DIM + e_{ijk} \qquad [1]$$

where Y_{ijk} is the phenotype of compositional and physical traits, $coag_i$ is the fixed effect of coagulation group (i=1, 2, 3), $herd_j$ is the fixed effect of herd (j = 1, 2, ..., 31), $parity_k$ is the fixed effect of parity (k= 1, 2, and >3), b_1 is a regression coefficient, *DIM* is a covariate and e_{ij} is the random residual effect. The general linear model was analysed with IBM SPSS Software (IBM Analytics).

For the fine-mapping of BTA18, a total of 395 animals genotyped for HDSNP were analysed, corresponding to the Swedish-Danish Milk Genomics Initiative (SLF-projects V0930028/V1130006). Single-SNP analyses were done using the animal model [2] in ASReml (Gilmour, et al. 2015):

$$NC_{ijklmo} = \mu + parity_i + herd_m + b_1 \cdot wim_{ijklmo} + b_2 \cdot e^{(-0.05 * wim_{ijklmo})} + b_3 \cdot CNcluster_o + b_4 \cdot SNP_k + animal_l + e_{ijklmo} \qquad [2]$$

where NC_{ijklm} is the phenotype; $parity_i$ is the fixed effect of parity (i=1, 2 and 3); $herd_m$ is the fixed effect of herd (m = 1, …, 21); wim_{ijklm} is the fixed effect of weeks in milk modelled by a Wilmink's curve (Wilmink, 1987); $CNcluster_o$ is the covariate describing the effect of the combined genotypes; SNP_k is the covariate representing 0, 1, or 2 copies of an allele; b_1-b_4 are regression coefficients; $animal_l$ is the random additive genetic effect assumed as N~(0, $\mathbf{G}\sigma_a^2$), where **G** is the genomic relationship matrix was built based on 395 cows with HDSNP genotypes, and σ_a^2 is the additive genetic, and $e_{ijklmo}$is the residual variance. Variance components were estimated with a model excluding the effect of SNP_k, and were fixed in model [2].

For the validation of a SNP in the VPS35 gene (Hylén, 2018), analyses of Real375, Imputed382, and Real582 were done in DMU (Madsen, et al. 2006), using the animal model [3]:

$$NC_{ijklm} = \mu + parity_i + herd_m + b_1 \cdot wim + b_2 \cdot e^{(-0.05 * wim)} + b_3 \cdot SNP + animal_l + e_{ijklm} \qquad [3]$$

where variables are as described previously for models [1 and 2], except for: for $animal_l$, assumed as N~(0, $\mathbf{G}\sigma_a^2$), where **G**-matrix was built based on 375 and/or 382 cows with HDSNP genotypes, or based on 582 cows with LDSNP genotypes. In addition, further comparisons were made between model [3] with and without the inclusion of the CN cluster. Variance components for model [3] with and without the inclusion of CNcluster were obtained in DMU, and heritability estimates were calculated as $h^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$.

For genomic prediction, two methods were compared: genomic best unbiased linear predictor (**GBLUP**) and Bayesian statistics (**MCMCglmm;** Hadfield, 2010). For both methods, a total of 982 animals were used, of which 382 Swedish RDC cows from the Swedish-Danish Milk genomics Initiative (SLF-projects V0930028/V1130006) and 600 Swedish RDC cows from the present project.

GBLUP estimates were obtained by running animal model [4] in DMU:

$$NC_{ijlm} = \mu + parity_i + b_1 \cdot wim_{ijlm} + b_2 \cdot e^{(-0.05*wim_{ijlm})} + animal_l + herd_m + e_{ijlm} \quad [4]$$

where variables are as described previously for models [1 and 2], except for the assumptions of the **G**-matrix. The G-Matrix was built based on 982 animals with MDSNP (genotyped and imputed) genotypes.

MCMCglmm estimates were obtained by running model [4] in R (R core team, 2018), with other assumptions than for GBLUP**.** Uninformative priors for fixed and random effects were used. The prior distribution applied to fixed effects was a normal distribution, while an inverse Gamma distribution (0.001; 0.001) was applied for each of the random effects. A MCMC chain ran for 1,200,000 iterations, after the initial burn-in of 30,000 iterations. Every 40th iteration was sampled for a total of 29,250 posterior samples. All variance components were estimated from the posterior distribution. All the posterior samples were averaged to obtain estimates of genomic breeding values (**GEBV**).

Cross-herd validation. The 982 Swedish RDC cows belong to 52 herds. Therefore, cross validation consisted in masking the phenotypes of each herd, and these masked herds, one at a time, were predicted with model [4] for GBLUP and for MCMCglmm. For each prediction, all other 51 herds were used as training population.

Prediction accuracies. To compare the prediction accuracies for each herd, a) correlation between the phenotype adjusted for fixed effects (**AdjPheno)** and GEBV divided by the square-root of the heritability, and b) receiver operating characteristics (ROC) curves were reported. While correlations were calculated in SAS 9.4, ROC curves were produced using the R-package pROC (Robin et al., 2011).

Furthermore, genetic correlations between NC milk (i) and other traits (j) measured in the present project were estimated from bivariate analyses of model [4] in ASReml. The correlation was calculated as: $Covariance(i,j)/\left(\sqrt{\sigma_i^2 * \sigma_j^2}\right)$.

Results and discussion

Some of the analyses are still ongoing at the time of writing this final report or under reviewer revision for international peer-reviewed scientific journals.

Rheological properties, milk composition and physical traits of the milk samples

The rheological measurements in the present study resulted in 18.1% of the milk samples being defined as NC and additional 18.9% of the milk samples as PC (Table 1). This results in a total of 37% of the milk from the Swedish RDC with impaired coagulation properties, which may influence industrial cheese production.

Table 1. Definition and frequency for the three coagulation groups as well as mean ± standard error of rennet coagulation time from the screening rheometer (RCT_{rox}).

	Non-coagulating milk (n=123)	**Poor coagulating milk (n=128)**	**Coagulating milk (n=428)**
Definition based on RCT_{rox}	≥ 40 min	16-39 min	≤ 15 min
Frequency (%)	18.1	18.9	63.0
RCT_{rox} (min)	44.79±0.42	21.64±0.58	10.68±0.10

Figure 1 shows the distribution of cows producing NC and PC milk across the 31 sampled farms. All farms have cows producing PC milk and 26 of the farms also have cows producing NC milk, which indicates that the problem is not caused by feeding or breeding on farm level but rather on national level.

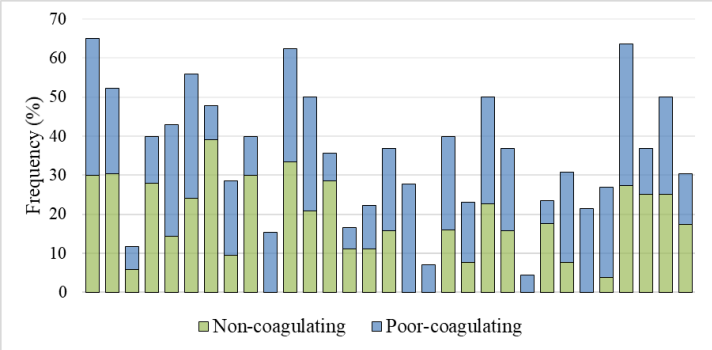

Figure 1. The frequency of cows producing non- and poor coagulating milk on the 31 studied farms.

The results from the rheological measurements, milk composition and physical traits for the three coagulation groups can be seen in Table 2. The results showed that NC milk can be linked to higher lactose content, less casein in relation to total protein, lower concentrations of both total calcium and Ca^{2+} as well as a higher titratable acidity, compared to coagulating milk. In addition, PC and NC milk seem to be caused by different mechanisms, as PC milk did not differ significantly for the same parameters as NC milk. The high occurrence of NC milk is believed to be caused by the focus of the breeding programs in the Nordic countries, where NC milk is unwantedly and indirectly breed for as an effect of breeding for a high protein content of the milk.

Table 2. Mean ± standard error of rheological parameters, milk composition traits and physical properties from Swedish Red Dairy Cattle based on different coagulation groups.

Trait	**Non-coagulating milk**	**Poor coagulating milk**	**Coagulating milk**
CFR_{rox} (Pa min^{-1})	$0.12±0.34^{a}$ (n=123)	$0.57±0.32^{a}$ (n=128)	$5.74±0.20^{b}$ (n=428)
G'_{max_rox} (Pa)	$2.60±9.28^{a}$ (n=123)	$97.88±8.78^{b}$ (n=128)	$266.53±5.32^{c}$ (n=428)
$G'_{40\ min\_stress}$ (Pa)	$9.20±6.09^{a}$ (n=52)	$89.43±5.43^{b}$ (n=67)	$172.84±3.39^{c}$ (n=205)
RCT_{stress} (min)	$25.39±0.71^{a}$ (n=52)	$8.95±0.63^{b}$ (n=67)	$5.40±0.39^{c}$ (n=205)
σ_{y_stress} (Pa)	$4.62±0.79^{a}$ (n=52)	$16.07±0.70^{b}$ (n=67)	$20.87±0.44^{c}$ (n=205)
Milk yield (kg day^{-1})	$31.7±0.62^{a}$ (n=123)	$33.3±0.59^{a}$ (n=128)	$32.1±0.36^{a}$ (n=428)
Somatic cell count (1000 cells mL^{-1})	$43.8±5.38^{a}$ (n=123)	$55.0±5.09^{a,b}$ (n=128)	$60.7±3.01^{b}$ (n=427)
Fat (%)	$4.37±0.15^{a}$ (n=123)	$4.40±0.15^{a}$ (n=128)	$4.49±0.09^{a}$ (n=427)
Lactose (%)	$4.66±0.02^{a}$ (n=123)	$4.66±0.02^{a}$ (n=128)	$4.59±0.01^{b}$ (n=427)
Protein (%)	$3.68±0.03^{a}$ (n=123)	$3.58±0.03^{b}$ (n=128)	$3.66±0.02^{a}$ (n=427)
Casein (%)	$2.87±0.03^{a}$ (n=119)	$2.79±0.02^{b}$ (n=127)	$2.87±0.01^{a}$ (n=406)
Casein:Protein	$0.77±0.00^{a}$ (n=119)	$0.77±0.00^{a,b}$ (n=127)	$0.78±0.00^{b}$ (n=406)
Calcium (mg kg^{-1})	$1168±17.3^{a}$ (n=123)	$1191±16.4^{a}$ (n=128)	$1280±9.9^{b}$ (n=427)
Ca^{2+} (mmol L^{-1})	$2.41±0.06^{a}$ (n=80)	$2.58±0.05^{b}$ (n=80)	$2.86±0.04^{c}$ (n=254)
Citric Acid (%)	$0.19±0.00^{a}$ (n=119)	$0.19±0.00^{a}$ (n=127)	$0.19±0.00^{a}$ (n=406)
pH	$6.69±0.01^{a}$ (n=123)	$6.68±0.01^{a}$ (n=128)	$6.70±0.00^{a}$ (n=428)
Casein micelle size (µm)	$0.142±0.00^{a}$ (n=119)	$0.140±0.00^{a}$ (n=116)	$0.139±0.00^{a}$ (n=374)
Titratable acidity (°N)	$21.8±0.77^{a}$ (n=16)	$20.0±0.92^{a,b}$ (n=9)	$19.4±0.50^{b}$ (n=29)

$^{a-c}$ Different superscript letters indicate significant differences ($P \leq 0.05$) within each row.

Pearson correlation coefficients were estimated to evaluate the correlation between NC milk and compositional and physical traits (Table 3). The calculated Pearson correlation coefficients showed that the binary NC trait was significantly correlated to all the rheological traits. NC milk could also be negatively correlated to fat content, casein:protein ratio, as well as both total calcium and Ca^{2+}. Moreover, NC milk was positively correlated to lactose content as well as titratable acidity. All the significant correlations between NC milk and the phenotypical traits were, however, weak (<0.4; Table 3). This can indicate that the traits measured in this study do not fully explain NC milk, but only to a certain degree, which means that none of the correlated parameters alone can be used to estimate NC milk behavior.

Table 3. Calculated Pearson correlation coefficients for rheological and phenotypical traits.

Parameter[1]	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1. NC																			
2. CFR_{rox} (Pa min^{-1})	**-0.38**																		
3. $RCT_{\_rox}$ (min)	**0.92**	**-0.55**																	
4. G'_{max_rox} (Pa)	**-0.61**	**0.88**	**-0.77**																
5. G'_{40_stress} (Pa)	**-0.70**	**0.87**	**-0.83**	**0.94**															
6. $RCT_{\_stress}$ (min)	**0.81**	**-0.42**	**0.82**	**-0.63**	**-0.69**														
7. σ_{y_stress} (Pa)	**-0.67**	**0.63**	**-0.69**	**0.73**	**0.81**	**-0.68**													
8. Milk yield (kg day^{-1})	0.03	**-0.29**	0.06	**-0.19**	**-0.25**	**0.11**	**-0.23**												
9. SCC (1000 cells mL^{-1})	-0.07	**0.14**	**-0.09**	**0.09**	0.03	-0.02	0.05	**-0.19**											
10. Fat (%)	**-0.09**	**0.21**	**-0.10**	**0.20**	**0.22**	0.07	**0.15**	**-0.46**	**0.18**										
11. Lactose (%)	**0.09**	**-0.24**	**0.13**	**-0.19**	**-0.26**	**0.11**	**-0.23**	**0.32**	**-0.39**	**-0.46**									
12. Protein (%)	-0.02	**0.50**	-0.06	**0.38**	**0.42**	-0.07	**0.44**	**-0.48**	**0.15**	**0.29**	**-0.35**								
13. CN (%)	-0.03	**0.49**	-0.08	**0.39**	**0.38**	-0.05	**0.41**	**-0.51**	**0.16**	**0.37**	**-0.32**	**0.93**							
14. CN:Protein	**-0.09**	**0.08**	**-0.11**	**0.13**	0.05	-0.05	0.08	**-0.08**	0.05	0.07	**-0.10**	0.03	**0.20**						
15. Calcium (mg kg^{-1})	**-0.20**	**0.41**	**-0.25**	**0.35**	**0.41**	**-0.20**	**0.38**	**-0.25**	-0.03	**0.14**	-0.06	**0.39**	**0.38**	0.01					
16. Ca^{2+} (mmol L^{-1})	**-0.22**	**0.22**	**-0.25**	**0.27**	**0.21**	**-0.23**	**0.22**	-0.10	0.05	**0.24**	**-0.38**	0.05	-0.01	**0.19**	-0.01				
17. Citric Acid (%)	-0.01	-0.02	-0.01	0.00	0.05	-0.05	0.00	0.02	**-0.11**	0.05	-0.00	**-0.16**	**-0.12**	0.05	**0.18**	**0.11**			
18. pH	0.00	**-0.09**	0.00	**-0.10**	-0.06	-0.02	**-0.13**	-0.01	**0.12**	-0.07	-0.02	**-0.14**	**-0.15**	0.06	-0.02	**-0.26**	0.07		
19. D[4,3] (nm)	0.00	-0.06	-0.02	-0.04	-0.07	-0.03	0.03	-0.07	0.06	**0.11**	**-0.24**	0.04	0.06	**0.17**	-0.03	**0.23**	0.00	**-0.10**	
20. TA (°N)	**0.33**	-0.12	**0.35**	-0.16	-0.10	-0.06	0.06	-0.27	0.09	0.25	-0.08	**0.40**	**0.40**	-0.16	-0.09	**-0.31**	0.21	**-0.42**	0.12

[1] NC = binary trait for non-coagulating milk samples. Bold numbers are significant ($P \leq 0.05$).

In order to evaluate where in the coagulation process the issue of NC milk is present, coagulating milk samples and NC milk samples before and after rennet addition were compared using capillary electrophoresis (Figure 2). It can be seen that the rennet is cleaving the κ-casein into para-κ-casein and caseinomacropeptide (CMP) in both coagulating and NC milk samples, which means that the mechanism of NC milk is linked to the aggregation step and not the enzymatic step of rennet-induced coagulation. This supports the lower Ca^{2+} content found in NC milk in the present study, since Ca^{2+} can increase the aggregation ability of rennet-induced coagulation.

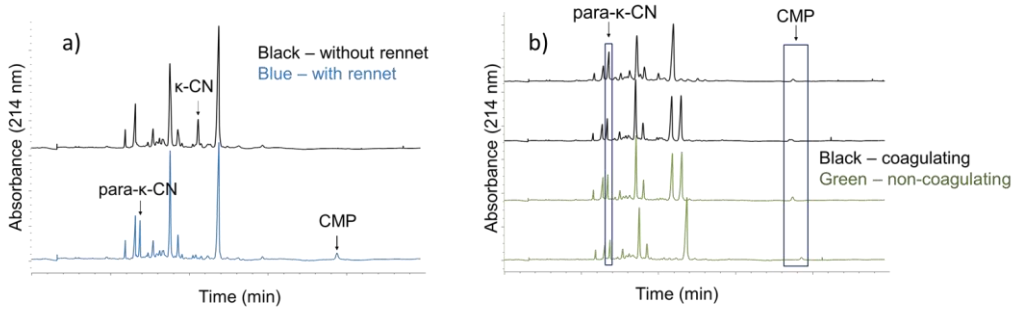

Figure 2. Electrograms from the capillary electrophoresis measurements showing a) the same sample before and after rennet addition and b) two coagulating and two NC samples after rennet addition. CMP=caseinomacropeptide.

Fine-mapping of *Bos Taurus* Autosome (BTA) 18

With imputed sequences, a genomic region on BTA18 of 30 mega base pairs was fine-mapped. For this region, a total of 205 variants were significantly associated with NC milk at −Log10 (*P*-value) > 6 (Duchemin et al., 2016). The most significant variants were one indel (rs385975260)

and two SNP (rs525335650 and rs379827811). These one indel and two SNP are in perfect LD with each other. (Figure 3A). After adjusting for the effect of rs525335650 (TagSNP1) as a fixed effect in model [2], a total of 80,206 variants were re-analyzed and no remaining associations were found (Figure 3B). This suggest that a QTL influencing NC milk occurs between 15,03-15,04 MBP. This QTL is in close proximity to the VPS35 gene.

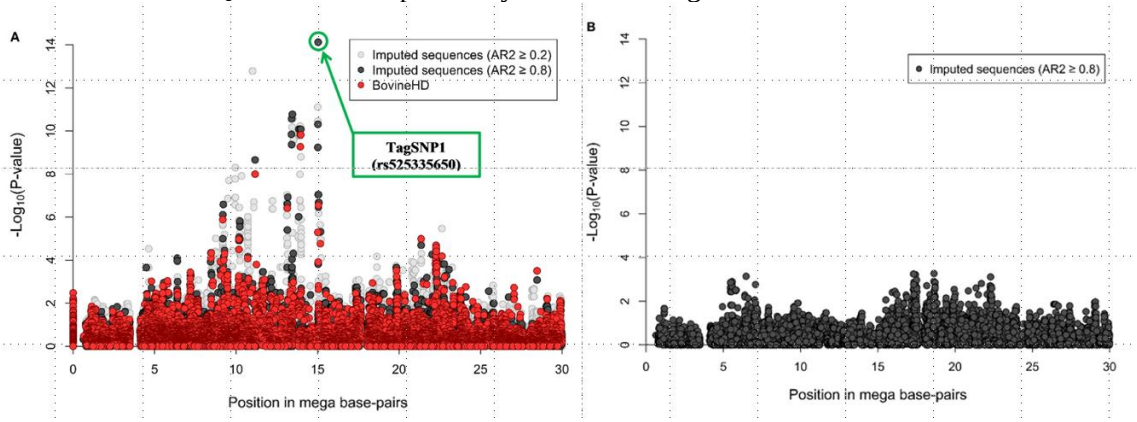

Figure 3. Fine-mapping of 30MBP on BTA 18. (A) of 137,949 polymorphic imputed variants overlaid with the HDSNP genotypes. In light gray, imputed variants with accuracy of imputation (AR2) ≤ 0.2. In black, imputed variants with AR2 ≥ 0.8. "TagSNP1" as most significant association. (B)After correcting for TagSNP1. In black, imputed variants with AR2 ≥ 0.8(N = 80,206 variants).

Validation of a SNP in the VPS35 gene

The validation of a SNP in the VPS35 gene started by comparisons between 3 different data sets: Real375, Imputed 382 and Real 582 (Table 4). Furthermore, comparisons between the variance components were similar between Real375 and Imputed382 for model [3] (Table 4). However, estimates for residual variance was higher in Real582 than for the other sets. The G-matrix for Real582 was built with LDSNP while the G-matrix of the other two data sets were built with HDSNP. This may partly explain the lower genetic variance and the higher residual variance seen in Real582. Another factor to take into account is the modelling of the herd effect as a fixed effect (Hylén, 2018).

Table 4. Means (SD) of NC milk across the three data sets, and measures of variability (heritability estimates (SE), phenotypic variance (SE) and residual variance) obtained with model [3] with and without CN cluster.

	Real375	Imputed382	Real582
NC milk[1]	0.18 (0.38)	0.18 (0.38)	0.19 (0.39)
h^2 _model[3]	0.39 (0.14)	0.34 (0.14)	0.17 (0.09)
Phenotypic variance_model [3]	0.13 (0.36)	0.12 (0.35)	0.15 (0.39)
Additive genetic variance_model [3]	0.05 (0.22)	0.04 (0.20)	0.03 (0.16)
Residual variance _ model [3]	0.08	0.08	0.12
h^2_model [3] + CN^2	0.47 (0.15)	0.34 (0.15)	
Phenotypic variance_model [3] + CN^2	0.12 (0.35)	0.11 (0.33)	
Additive genetic variance_model [3] + CN^2	0.06 (0.24)	0.04 (0.19)	
Residual variance _ model [3] + CN^2	0.06	0.07	

[1]Scored as a binary trait; [2] CN= CNcluster

The allele frequencies of the SNP in the VPS35 gene were similar between the three sets (Table 5). Although Imputed382 and Real375 differed in genotypes by seven observations, the correlation between the genotypes in the Imputed382 and the Real375 was 0.77 (Hylén, 2018). For the estimated effects of genotypes (Table 5), estimates from model [3] suggest that the C allele shows additive behavior with effects in the same direction for both Real375 and Imputed382, although an underestimation of the Imputed382 is noticeable too. For Real582, GG and GC estimates are similar suggesting that the effects could not be disentangled from one

another. This reflects the lower genetic variance estimated for Real582 in model [3]. The inconsistent result of cows from the Swedish-Danish Milk Genomics Initiative (SLF-projects V0930028/V1130006) compared with cows from the present project may indicate that the results from the earlier study by Duchemin *et al.,* (2016) is a false positive and that the VPS35 SNP genotyped may not be a causal variant for NC milk. Nevertheless, there is also a possibility of sample contamination during the DNA extraction and PCR. Estimates from model [3] with CN cluster suggest that Real375 and Imputed 382 results are in the same direction, however for GG, the estimates seem to be more similar between the two sets. Differences between the estimated effects of GG genotypes between model [3] with and without CN cluster might not be caused by differences in the allele frequencies between the two data sets (Hylén, 2018).

Table 5. Distribution of genotypes across cows, allele frequencies, and estimated effects of the SNP in the VPS35 gene for models [3 and 4].

VPS35 SNP	Real375			Imputed382			Real582	
	N	Effect model[3]	Effect model[4]	N	Effect model[3]	Effect model[4]	N	Effect model[3]
CC	2	0.00	0.00	1	0.00	0.00	2	0.00
GC	23	-0.58 (0.26)	-0.53 (0.25)	21	-0.28 (0.36)	-0.30 (0.35)	44	0.38 (0.29)
GG	350	-1.01 (0.25)	-0.96 (0.25)	360	-0.88 (0.35)	-0.92 (0.34)	536	0.24 (0.28)
F(G)	0.96			0.97			0.96	
F(C)	0.04			0.03			0.04	

Nevertheless, genetic correlations estimated between NC milk and other traits (Table 6) for the present project suggest that other regions of the genome or genes may be affecting NC milk. This is reflected by the strong genetic correlations with Ca^{2+}, β-CN, α_{s1}-CN and α_{s2}-CN. Furthermore, strong genetic correlations ranging from -0.98 (G'_{max_rox}) to -0.76 (σ_{y_stress}) suggest that NC milk is strongly correlated with rennet-induced coagulation properties.

Table 6. Genetic correlations (SE) between NC milk and other traits based on 724 cows from the present project.

Trait	**Fat (%)**	**Protein (%)**	**CRF_{rox} (Pa min^{-1})**	**G'_{max_rox} (Pa)**	**σ_{y_stress} (Pa)**
NC milk	-0.20(0.30)	0.12(0.26)	-0.83(0.17)	-0.98(0.09)	-0.76(0.26)
	Ca^{2+} (mmol L^{-1})	**β-CN (wt/wt%)**	**κ-CN (wt/wt%)**	**α_{s1}-CN (wt/wt%)**	**α_{s2}-CN (wt/wt%)**
NC milk	-0.65(0.23)	-0.83(0.21)	-0.02(0.28)	0.53(0.19)	0.66(0.18)

Genomic selection against NC milk

The cross-herd validation was still ongoing at the time of writing this final report, thus only the results for GBLUP were available and are discussed.

Prediction accuracies from the cross-herd validation

For GBLUP, positive prediction accuracies show that herds 36 (r=1.22) and 17 (r=1.11) were very well predicted by GBLUP (Figure 4). The composition of the validation data set was of: a) 6 cows (1 cow classified as NC milk) in herd 36, and b) 12 cows (1 cow classified as NC milk) in herd 17. Although the overall prediction average was r=0.30 (Figure 4), there were negative prediction accuracies for herds 48 (r=-0.50), 3 (r=-0.52), 43 (r=-0.51) and 19 (r=-0.50). The composition of the validation data set was of: 22 cows (5 cows classified as NC milk) for herd 48; b) 22 cows (0 cows classified as NC milk) in herd 3; 12 cows (1 cow classified as NC milk) for herd 43; b) 16 cows (2 cows classified as NC milk) in herd 19. It is possible that the predicted accuracies are influenced by the frequency of NC milk in the herds.

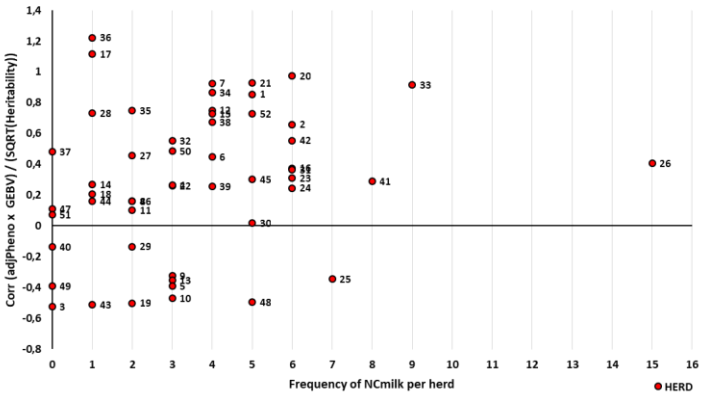

Figure 4 – prediction accuracy for GBLUP calculated as a correlation between adjusted phenotype and GEBV divided by the square root of the heritability. The training set was of 51 herd, and the validation was of 1 herd at a time. Results here are for the validation data set, for which correlations are plotted against the frequency of NC milk per herd.

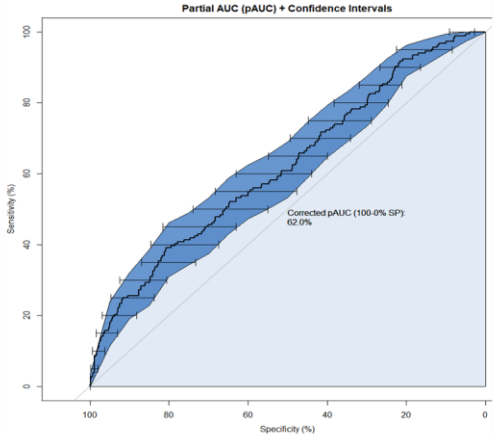

Figure 5 – prediction accuracy for GBLUP calculated as ROC curves. The training set was of 51 herd, and the validation was of 1 herd at a time. Results here are for the validation data set.

An ROC plot shows the performance of a binary classification method with discrete ordinal output (Figure 5), by showing the proportion of correct classified positive observations (sensitivity) and the proportion of correct classified negative observations (specificity). The closer the ROC curve is to the upper left corner, the more efficient the predictions are. For GBLUP, the ROC curve (Figure 5) and the partial area under the ROC curve (pAUC) is equal to 0.62. This suggest that our GBLUP predictions range from poor to fair according to the CLSI/NCCLS protocols (2008). For MCMCglmm, a better performance as compared with GBLUP regarding correlations and ROC curves is expected. These results suggest that genomic selection could be a tool to reduce the prevalence of NC milk.

Conclusions

By studying a large number of cows, the project has confirmed that 18.9% of milk from Swedish RDC is poorly coagulating and as much as 18.1% is non-coagulating, which is an alarming high figure of 37% of the cows that give milk that has difficulties to coagulate. This indicates that milk from the Swedish RDC has an alarming high prevalence of milk with impaired coagulation properties, which may influence industrial cheese production. All the farms sampled in this study had cows producing PC milk and most of them also NC milk. By this, it can be concluded that the impaired milk is not just a problem on farm level but rather on a national level. An increased understanding of milk composition and physical traits affecting NC milk has been achieved, specifically a low calcium content and aggregation stage of the coagulation process, and can be used to further improve the processing quality of milk. The project has identified selectable markers and putative mutations for NC milk, thereby suggesting a link from mutation to gene expression and subsequent effects on coagulation. Furthermore, markers and mutations associated with NC milk have been identified in the Swedish-Danish Milk Genomics Initiative sample set (SLF-projects V0930028/V1130006), but could not be confirmed in the new samples in the present project. The level of the prediction accuracy suggests that there is scope to improve this trait through genomic selection. Strong genetic correlations were estimated between NC milk and milk coagulation properties suggesting that NC milk can be modified by selective breeding, without negatively changing other nutritional and processing properties of the milk.

Benefits for stakeholders and industry as well as recommendations

An increased understanding of milk composition and physical traits affecting NC milk has been achieved, specifically a low calcium content and aggregation stage of the coagulation process, and can be used to further improve the processing quality of milk. The outcome of this project

could be used to develop a genomic selection strategy for NC milk, which can be used to restrict the use of the bulls that convey an increased risk of NC milk. For direct selection on individual haplotypes or genes, further validation studies of larger population samples are required. This could be applied in the Nordic breeding program and thereby ensure a potential benefit for the Swedish dairy and agricultural industries as well as stakeholders. By including coagulation ability in the breeding objective, the trait can be improved by selection and the problem of NC milk could be reduced. In addition, a homogeneous milk with minimal variation in quality can be delivered to the Swedish dairies. With an improved selection, an increased genetic progress for milk quality traits can be achieved. The dairy processing industry can differentiate milk collection and processing or provide financial incentive to the farmers. It is highly recommended to develop techniques for routine measurement of NC milk as well as to further explore the relations between NC milk and other nutritional and processing properties of the milk. Further exploration of the use of MIR spectra to predict NC milk could be one avenue to pursue, another is to develop routine measures of NC biomarkers for on-farm application.

References

CLSI/NCCLS document EP12-A2 User Protocol for Evaluation of Qualitative Test Performance; Approved Guideline 2nd Edition. Vol. 28 No. 3. 2008.
Daetwyler et al. 2014. Nat Genet 46, 858–865.
Duchemin et al. 2016. Front Genet 7:57.
Gilmour, et al. 2015. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.
Gregersen et al. 2015. J. Dairy Sci. 98:1261–1272.
Gustavsson et al. 2014a. Int. Dairy J. 39:102–107.
Gustavsson et al. 2014b. J. Dairy Sci. 97:5219–5229.
Hadfield, J. D. 2010. J Stat Softw, 33(2), 1-22 http://www.jstatsoft.org/v33/i02/.
Hylén, G. 2018. M.Sc. Thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden.
Madsen, et al. 2006. In: 8th WCGALP, Belo Horizonte, Brazil.
R Core Team 2018. R Foundation for Statistical Computing, Vienna, Austria.
Robin et al. 2011. BMC Bioinformatics 12:77.
SLF-project V0930028.
SLF-project V1130006.
Wilmink, J. B. M. 1987. Livest Prod Sci, 16(4), 335-348.

Part 3: Dissemination of results

Peer-reviewed scientific publications	**Published peer-reviewed scientific publications:**
	Duchemin S.I, Glantz M., De Koning D.J., Paulsson M. and Fikse W.F. 2016. Identification of QTL on chromosome 18 associated with non-coagulating milk in Swedish Red cows. *Frontiers in Genetics* 7:57. DOI: 10.3389/fgene.2016.00057
	Submitted manuscripts:
	Nilsson K., Stålhammar H., Stenholdt Hansen M., Lindmark-Månsson H., Duchemin S.I., Fikse F., de Koning D-J., Paulsson M. and Glantz M. 2018. Characterisation of non-coagulating milk and effects of milk composition and physical properties on rennet-induced coagulation in Swedish Red Dairy Cattle. *Under revision.*
	Planned publications/manuscripts:
	Nilsson K., Abdelghani A., Paulsson M. and Glantz M. Determination of para-kappa-casein in non-coagulating milk using capillary electrophoresis. *Manuscript in preparation.*
	Nilsson K., Buhelt Johansen L., Stålhammar H., Stenholdt Hansen M., Lindmark-Månsson H., Duchemin S.I., Fikse W.F., de Koning D-J., Paulsson M. and Glantz M. Effect of milk protein variants on rennet coagulation in milk from Swedish Red cows. *Manuscript in preparation.*
	Duchemin S.I., Nilsson K., Stålhammar H., Stenholdt Hansen M., Lindmark-Månsson H., Fikse W.F., de Koning D-J., Paulsson M. and Glantz M. Genetic parameters and rennet coagulation properties in milk from Swedish Red cows. *Manuscript in preparation.*
	Duchemin S.I., Gregersen V. R., Glantz M., Paulsson M., Pinto F.A.L., Fikse W.F. and de Koning D-J. Genetically differentiated variants between Swedish Red cows with non-coagulating and well-coagulating milk. *Manuscript in preparation.*
	Duchemin S.I., Tesfaye Y.G., Gustavsson F., Johansson A.M., Glantz M., Paulsson M., Fikse W.F. and de Koning D-J. Genome-wide association study of five milk composition traits in Swedish Red cows. *Manuscript in preparation.*
	Duchemin S.I., Nilsson K., Glantz M., Paulsson M., Fikse W.F. and de Koning D-J. Genomic selection against NC milk in Swedish Red cows assessed by cross-validation. *Manuscript in preparation.*
	Duchemin S.I., Nilsson K., Glantz M., Paulsson M., Fikse W.F. and de Koning D-J. Fine-mapping of 4 BTA with imputed sequences influencing NC milk in Swedish Red cows. *Manuscript in preparation.*

Other publications (Popular science publications etc)	Nilsson K., Duchemin S.I, Fikse W.F., Stenholdt-Hansen M., Lindmark-Månsson H., Stålhammar H., de Koning D-J., Paulsson M. and Glantz M., ” Avla kor i en ostligare riktning.” *Planned popular science article in industry/stakeholder magazine*, Spring 2019.
	De Koning D.J., Duchemin S., Glantz M. and Paulsson M., ”Jakt på mutation som stör osttillverkningen.” *SLU newsletter*, March 9, 2017.
	Paulsson M., Glantz M., Hedlund M., Nilsson K., Andersson I-M. and Edén J., ”Mejeriteknologi vid Lunds Universitet.” *Mjölk Special*, January 13, 2017.
Oral communication	Duchemin S.I, Glantz M., De Koning D.J., Paulsson M. and Fikse W.F., “Fine-mapping of a QTL region on BTA18 affecting non-coagulating milk in Swedish Red cows.” *66th Annual Meeting of the European Federation of Animal Science (EAAP), Warsaw, Poland,* August 31 – September 4, 2015.
	Glantz M., ”Vad betyder kons gener för mjölk, ost och yoghurt?” *Aktuell forskning och utbildning vid Lunds Universitet, Mejeritekniskt Forum, Lund, Sweden*; February 4, 2016.
	Glantz M. and Duchemin, S.I., “Genetic aspects of non-coagulating milk.” *Nordic Workshop in Dairy Cattle Genomics 2017, Copenhagen, Denmark*, April 27, 2017.
	Nilsson K., Duchemin S.I, Fikse W.F., Stenholdt-Hansen M., Lindmark-Månsson H., Stålhammar H., de Koning D-J., Paulsson M. and Glantz M., “Genomic selection against non-coagulating milk in Swedish Red cows.” *Nordic Dairy Congress 2017, Copenhagen, Denmark,* June 7-9, 2017.
	Nilsson K. “Genomic selection against non-coagulating milk.” *Workshop within Nordic-Baltic Dairy Network, Larkollen, Norway,* November 20-22, 2017.
	Duchemin S.I., Gregersen V., Glantz M., Paulsson M., Pinto F., Fikse F. and de Koning D.J., “Variants on BTA12 are genetically differentiated between Swedish Red cows with non- and well- coagulating milk.” *11th World Congress on Genetics Applied to Livestock Production (WCGALP) 2018, Auckland, New Zealand,* February 7-11, 2018.
	Duchemin S.I., Gregersen V., Glantz M., Paulsson M., Pinto F., Fikse F. and de Koning D.J. “Variants on BTA12 are genetically differentiated between Swedish Red cows with non- and well- coagulating milk.” *Nordic Workshop in Dairy Cattle Genomics 2018, Copenhagen, Denmark,* April 17-18, 2018.
	Paulsson M. and Glantz M., “Dairy Technology at Lund University.” *Steering Committee meeting for Education and Research in Dairy Technology at Lund University, Lund, Sweden,* December 1, 2015; June 2, 2016; December 1, 2016; June 1, 2017; December 1, 2017; June 1, 2018; November 30, 2018.
	Glantz M., Nilsson K., Duchemin S.I., Paulsson M., Fikse F., Stålhammar H., Stenholdt Hansen M., Lindmark-Månsson H. and de Koning D-J. “Sum-up of SLF-project O-15-20-274: Genomic selection against non-coagulating milk.” *Planned workshop for industry and stakeholders, Lund/Uppsala, Sweden,* Autumn 2019.

Student theses	Hylén G. "Validation of the role of a SNP in the VPS35 gene and its effect on non-coagulating milk in Swedish Red Cattle", Degree project, Swedish University of Agricultural Sciences, Sweden, 2018.
	Bender P.L. "Prediction of milk coagulation - Prediction and classification of coagulation properties by Fourier transform infrared spectroscopy and chemometrics", Bachelor project, Copenhagen University, Denmark, 2018.
	Carlsson, M. "Characterization of non-coagulating milk – Compositional, physical, and structural properties in comparison to coagulating milk", MSc. degree project, Lund University, Sweden, 2017.
Other (posters)	Duchemin S.I, Glantz M., De Koning D.J., Paulsson M. and Fikse W.F., "Fine-mapping of a QTL region on BTA18 affecting non-coagulating milk in Swedish Red cows." *Poster at International Dairy Federation Dairy Science & Technology Symposia 2016, Dublin, Ireland.*
	Nilsson K., Duchemin S., Fikse W.F., Stenholdt-Hansen M., Lindmark-Månsson H., Stålhammar H., de Koning D-J., Paulsson M. and Glantz M., "Genomic selection against non-coagulating milk in Swedish Red cows." *Poster at International Dairy Federation World Dairy Summit 2016, Rotterdam, the Netherlands.*