UNIVERSITÀ
DEGLI STUDI
DI PADOVA

GENETICS OF MILK PROTEIN COMPOSITION AND MILK COLOUR IN IRISH DAIRY CATTLE

SEDE AMMINISTRATIVA: UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI AGRONOMIA ANIMALI ALIMENTI RISORSE NATURALI E AMBIENTE

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE ANIMALI E AGROALIMENTARI INDIRIZZO: PRODUZIONI AGROALIMENTARI CICLO XXVIII

TEAGASC ANIMAL & GRASSLAND RESEARCH AND INNOVATION CENTRE, MOOREPARK, FERMOY, CO. CORK, IRELAND

Direttore della Scuola: Ch.ma Prof.ssa Viviana Corich

Supervisore: Ch.mo Prof. Massimo De Marchi

Supervisore: Ch.mo Dott. Sinead McParland

Dottorando: Audrey McDermott

Table of Contents

Declaration

I declare that this thesis has not previously been submitted as an exercise for a degree at the University of Padova, or any other university, and I further declare that the work embodied in it is my own.

_____

Acknowledgements

I would like to acknowledge the opportunity presented to me by Teagasc and the University of Padova to carry out this work in the Animal and Grassland Research and Innovation Centre at Moorepark. I would also like to acknowledge the Teagasc Walsh Fellowship Scheme for funding the research.

I convey gratefulness to Dr Massimo De Marchi for his constant support and guidance over the past few years. I appreciate his contributions of time, ideas, invaluable advice and feedback; to make my PhD experience productive and stimulating.

I would like to thank my fellow office mates; Jessica, Alan and Tom for their support, reassurance, inspiration and feedback. To my great housemates; Deirdre, Aideen, Michelle and Grainne, even during the challenging days it was always a pleasure to return to the house in the evening to find you always ready to listen with chocolate and wine in hand.

I would like to thank my parents for all your love and encouragement, thank you for always being there to listen to me complain and give out 'when the going got tough'. It is a testament to you that I have gotten this far!

I dedicate my thesis to my beloved cousins Clodagh, Tadhg, Liam, Niall and Ryan may you rest in peace.

List of tables

List of figures

Glossary of terms

a*	Redness
AA	Amino Acids
α-LA	Alpha lactabumin
Arg	Arginine
αS2-CN	Alpha $_{S2}$ Casein
αS1-CN	Alpha $_{S1}$ Casein
Asp	Aspartic acid
b*	Yellowness
β-CN	Beta casein
β-Lg-A	Beta lactoglobulin A
β-Lg-B	Beta lactoglobulin B
CN	Casein
CV	Coefficient of variation
FAA	Free amino acids
Glu	Glutamic acid
Gly	Glycine
h^2	Heritability
IR	Infrared
NIRS	Near infrared spectroscopy
k-CN	Kappa casein
L*	Lightness
Lys	Lysine
MIRS	Mid-infrared
R	Coefficient of correlation
R^2	Coefficient of determination
Ser	Serine
SCC	Soamtic cell count
Total CN	Total casein
Total LG	Total lactoglobulin
Val	Valine
σ	Standard deviation

Thesis abstract

The overall aim of this thesis was to determine the feasibility of breeding for improved milk quality and in particular protein fractions, free amino acids (FAA) and milk colour. To breed for a characteristic such as milk quality it must be; (i) economically or socially important (ii) exhibit genetic variation (i.e be heritable), and (iii) be measurable or genetically correlated with a measurable trait. Gold standard data was determined from 715 milk samples. Spectral data used consisted of ~ 95,000 milk samples from seven research herds and ~ 40,000 milk samples (morning and evening milk samples combined) from 69 commercial herds. The greatest correlation coefficients of external validation obtained for protein fractions, FAA and milk colour were 0.74 (total casein), 0.75 (glycine) and 0.72 (yellowness), respectively. Milk protein fractions and FAA change across calendar months of the year, stage of lactation and parity. A peak in the concentration of all casein fractions was evident in the months of August, September and October. The concentration of glutamic acid was greatest during the months of February, March, April and June when adjusted for milk yield. Changes in individual milk protein fractions and FAA across calendar months of the year and across stages of lactation could provide useful input parameters for decision support tools in the management of product portfolios by processors over time. Heritability of the predicted protein fractions and FAA ranged from 0.04 (beta casein) to 0.61 (total lactoglobulin) and from 0.05 (aspartic acid) to 0.58 (serine), respectively. The coefficient of genetic variation of gold standard protein fractions and FAA ranged from 3.01 (alpha lactalbumin) to 22.98 (total lactoglobulin) and from 1.01 (glutamic acid) to 25.65 (serine), respectively. Milk colour traits were low to moderately heritable ranging from 0.29 (lightness) to 0.35 (yellowness), respectively. The coefficient of genetic variation of milk colour ranged from 0.37 (lightness) to 1.72 (greeness), respectively. Results from this thesis clearly show that some protein fractions, some FAA and milk colour are predictable from MIRS and these predictions exhibit genetic variation and thus breeding for improved milk quality is feasible. The outcome of this thesis is primarily that the prediction of these traits by MIRS could benefit the dairy breeding industry worldwide through genetic selection of animals with higher quality milk and allowing for the more accurate selection of milk for human consumption, infant milk formula, and cheese production. The generated predictions could also be useful for herd and processor management strategies.

Riassunto della tesi

L'obiettivo generale della presente tesi è stato quello di determinare la possibilità di poter migliorare, tramite in programmi di selezione genetica, la qualità del latte e in particolare le frazioni proteiche, gli amino acidi liberi (FAA) e il colore. Per essere migliorato geneticamente un carattere (incluso la qualità del latte) deve: i) essere di importanza, sia essa economica o anche sociale; ii) esibire variabilità genetica, ossia deve essere ereditabile; iii) essere misurabile o correlato geneticamente con un carattere che sia misurabile. Le analisi di riferimento per i suddetti parametri di qualità del latte sono state determinate su 715 campioni di latte. Il dataset di spettri includeva misurazioni infrarosse su circa 95 000 campioni di latte raccolti in sette aziende sperimentali, mentre altri circa 40 000 spettri (determinati su campioni di latte di entrambe le mungiture giornaliere) erano provenienti da 69 aziende commerciali. I più alti coefficienti di correlazione, in validazione esterna, ottenuti per frazioni proteiche, FAA e colore del latte sono stati rispettivamente di 0.74 (caseine totali), 0.74 (glicina) e 0.72 (indice del giallo). Le frazioni proteiche del latte e gli FFA hanno dimostrato variazioni tra mesi dell'anno, tra stadi di lattazione e tra ordini di parto. Un picco nella concentrazione di tutte le frazioni caseiniche è stato evidente nei mesi di Agosto, Settembre ed Ottobre. La concentrazione di acido glutammico è stata maggiore nei mesi di Febbraio, Marzo, Aprile e Giugno a parità di produzione di latte giornaliera. Le variazioni di frazioni proteiche e FAA attraverso mesi dell'anno e stadi di lattazione possono fornire all'industria di trasformazione lattiero-casearia uno strumento per gestire il proprio portafoglio prodotti lungo uno specifico periodo produttivo. I valori di ereditabilità dei fenotipi predetti hanno avuto un minimo di 0.04 (beta caseina) ed un massimo di 0.61 (lattoglobulina totale) per le frazioni proteiche, mentre per quanto riguarda gli FAA hanno variato tra 0.05 (acido aspartico) e 0.58 (serina). Il coefficiente di variazione genetico per frazioni proteiche misurate ha variato tra 3.01% (alfa lattoalbumina) e 22.98% (lattoglobulina totale), mentre per gli FFA misurati ha variato tra 1.01% (acido glutammico) e 25.65% (serina). Il caratteri di colore del latte hanno dimostrato una ereditabilità medio-bassa, con un range compreso tra 0.29 (luminosità) e 0.35 (indice del giallo). Il coefficiente di variazione genetico del colore del latte ha avuto un minimo di 0.37% (luminosità) ad un massimo di 6.68% (indice del giallo). I risultati della presente tesi dimostrano chiaramente che alcune frazioni proteiche, alcuni FAA e il colore del latte sono di possibile predizione attraverso la tecnologia nel medio-infrarosso, e tali fenotipi predetti

hanno variabilità genetica il che implica che programmi di selezione per migliorare la qualità del latte sono possibili. I risultati principali di questa tesi sono che le predizioni di questi caratteri usando la spettroscopia nel medio infrarosso possono rappresentare un beneficio per gli allevatori di vacche da latte attraverso la selezione genetica di animali con una migliore qualità del latte. Inoltre, questa tesi offre delle opportunità per una selezione più accurata del latte destinato al consumo umano, alla produzione di latte per neonati e alla produzione di formaggio. Inoltre, tali predizioni possono rappresentare dlle opportunità per il management aziendale e industriale.

CHAPTER 1

Introduction

1.1 Infrared spectroscopy

Electromagnetic radiation is a form of energy that is propagated through free space or through a material medium in the form of electromagnetic waves. As shown in **Figure 1.1** different types of electromagnetic waves exist which differ in wavelength. The wavelength is the distance between one wave crest to the next. From the shortest wavelength to the longest, they are classified as following:

1) Gamma rays
2) X-rays
3) Ultraviolet visible light
4) Infrared, which is further categorised into:
    a. Near infrared waves
    b. Mid-infrared waves (short, medium and long waves)
    c. Far infrared waves
5) Microwaves
6) Radio waves

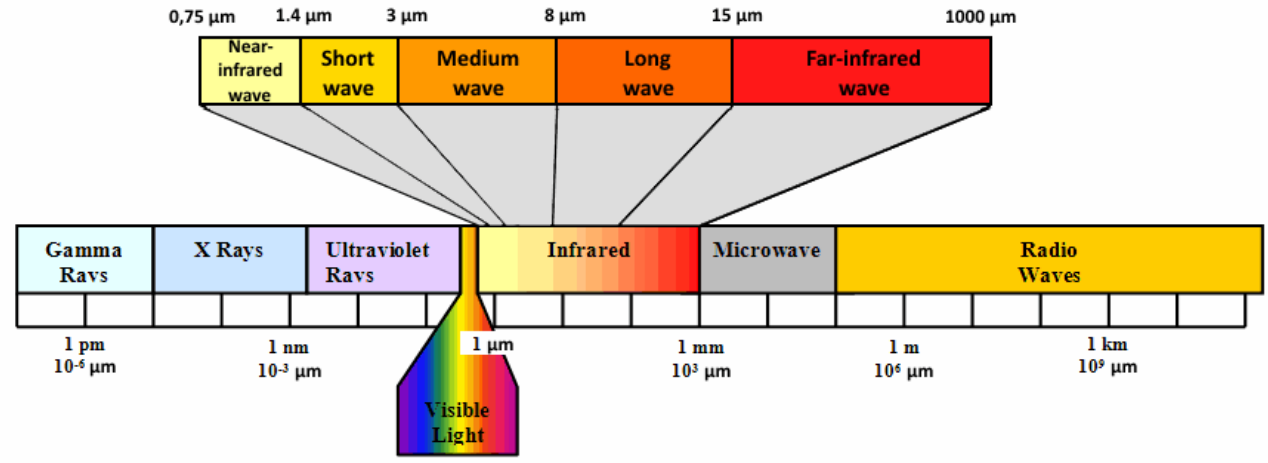


Figure 1.1 Regions of the electromagnetic spectrum (Pellizon Birelli and Fazio, 2005).

Infrared (IR) spectroscopy is the analysis of infrared light interacting with a molecule and it is based on the capability of every molecule to reflect, transmit or absorb part of the energy when this light hits it. The amount of energy absorbed (absorbance) is directly proportional to the amount of the absorbent molecule in the sample (Lambert-Beer law).

1.1.1 Advantages and disadvantages of infrared spectroscopy

Some of the advantages of IR spectroscopy technology (Williams, 2007) include that:

1) It is quick and efficient, yielding a response in real-time
2) It can study samples in almost any state, including liquids, solutions, pastes, powders, films, fibres, gases, and surfaces
3) It is precise and accurate (accuracy is equivalent to primary reference methods)
4) It is practical
5) It is inexpensive to operate (low labour requirements)
6) It is environmentally friendly
7) It is durable (approximately 10-year lifespan)
8) It is simple and safe to operate
9) It facilitates simultaneous analysis of several traits
10) It does not destroy the sample during the analysis

Some of the disadvantages of IR technology (Williams, 2007) include that:

1) Separate calibration is required for each product and constituent
2) Accuracy and reproducibility must be monitored
3) Equipment can be expensive to purchase
4) Training is required to operate instruments most efficiently

One of the complications of IR analyses of solutions is the tendency for common solvents such as water to strongly absorb infrared light therefore adding noise to the analyses. Noise is defined as any electronic signal that reaches the spectral detector that is not directly related to the actual absorption bands required for analysis. Water is a particularly poor solvent for use in the infrared region, as the spectral bands associated with O-H vibrations are very strong and broad. Essentially, water is able to "blank out" large regions of the spectrum, deleting important information and reducing accuracy. This is a disadvantage, as many biological, forensic, and clinical samples occur naturally in aqueous solutions; for example, bovine milk is 87.7% water. To deal with that inconvenience spectra can be manipulated to diminish noise by a smoothing process, which helps with the spectra interpretation.

1.1.2 History and development of infrared spectroscopy

One of the most significant historical events for IR spectroscopy was the development of the Fourier Transform in the 1700s; prior to this mathematical transformation was enhanced by the use of interferometer (e.g., Herschel and Michelson in 1800; Coblentz, Wright and Hersher, Barer, Fellgett, Jacquinot, Cooley, and Tuckey in the first half of 1900). The first Fourier Transform infrared spectrophotometer with a dedicated minicomputeron was sold in 1969 by Digilab, and was later modified in 1983, by the same company (Spectra-Tech). From the 1980s onwards, Fourier Transform infrared spectrophotometers were amalgamated with personal computers and this technique of analysis became popular because of its efficiency and cost-effectiveness.

Infrared spectroscopy is presently used in a wide range of scientific fields such as biology (Chang et al., 2001) and agriculture (Williams et al., 2007). Fourier Transform spectrometers are commonly used in chemistry to analyse organic and inorganic molecules, as well as polymers (Siesler et al., 2008) and in agriculture, animal production and food science to analyse meat quality, milk composition, forage, manure, and wine (Williams et al., 2007).

1.2 Near-infrared spectroscopy

Near-infrared spectroscopy (NIRS) is based on the absorption of electromagnetic radiation in the region from 750 to 1400 nm. The analysis of water by NIRS was the first successful application of this rapid technology, which has developed over the past 30 years into a routine method for many agricultural commodities and food constituents. In agriculture, NIRS is used to analyse soil (Cecillon et al., 2009), forage (Cozzolino and Moron, 2004), manure (Reeves and Van Kessel, 2000), wine (Cozzolino et al., 2008), milk (Albanell et al., 2003) and animal meat (Mitsumoto et al., 1991) composition. Therefore, agricultural applications of NIRS include farm management (Kawasaki et al., 2008), plant breeding (Cozzolino et al., 2000), flour milling (Osborne et al., 2007), and grain handling (Mohan et al., 2005). Other applications of NIRS are found in medical, biomedical studies, food science, forestry, pharmaceutical and petroleum industries.

1.3 Mid-infrared spectroscopy

Mid-infrared spectroscopy (MIRS) is a technique that studies the interactions between light and matter within the mid-infrared region of the electromagnetic spectrum (1,400-15,000 nm). Mid-infrared spectroscopy is widely used in the pharmaceutical

industry in chemical structural confirmation, to identify suspect counterfeit samples, and for the identification of active pharmaceutical ingredients in drug products (Dziki and Doddi, 2008). Agricultural applications of MIRS include quantification of product composition (including bovine milk; Dousseau et al., 1989), sugar, sugarcane, and beetroot (Mehrotra and Siesler, 2003). Mid-infrared spectroscopy is routinely used by milk recording organisations worldwide to quantify milk protein, casein (CN), fat and lactose content and has practical application both in milk payment schemes and in monitoring individual animal performance to aid animal management and breeding decisions.

1.3 Milk quality

Milk quality parameters include somatic cell count (SCC), total bacterial count, fat content and composition, mineral content and composition, protein content and composition, free amino acid (FAA) composition and milk colour . Milk quality influences the manufacturing process, yield, and consistency, affecting profit margins and market access. It is known that milk protein composition is important as it affects both yield and characteristics of cheese and plays a vital role in the production of all cheese types (De Marchi et al., 2009a).

Some of the desired milk quality parameters depends on the target market; for example, Irish consumers prefer a more yellower creamier milk with a higher fat content, whereas a yellower colour milk are considered an unfavourable attribute in Middle Eastern dairy markets (Keen and Wilson, 1992). The milk payment scheme in the majority of countries is partially based on specific milk quality parameters. Tiered milk payment penalises dairy producers for poor quality milk (high milk SCC and total bacterial) and it is now becoming increasingly common for producers to be financially rewarded on a range of different milk quality parameters including fat and protein content. Thus, the provision of premium quality milk is in the interest of all sectors.

Milk production in Ireland is generally grass based and therefore seasonal, with a peak: trough ratio of total milk yield of 7:1 (May vs. January). However, demand for liquid milk and other dairy products is constant year round (Department of Agriculture and Food, CSO, 2002). Ireland's grass based production system enables the production of high quality milk, as grass has a high energy and protein content (Teagasc, 2014). Currently, Ireland produces about 5.5 billion litres of milk per year and exports 85% of its liquid milk and milk products (ICOS, 2016). Given the recent abolishment of the milk

quota, there is an expected increase in milk supply of 50% by 2020 (Food Harvest, 2020). Marketing Irish milk as high quality milk produced on a grass-based system could give Ireland a competitive advantage over other countries. However, this requires detailed knowledge on the milk quality of the national herd, which is not feasible using currently available milk quality data.

Infant formula production is the fastest growing sector in the world dairy market (FAOSTAT, 2014) and the international market for infant milk formula is worth approximately US$5-6 billion annually. Ireland is the largest exporter of infant formula in Europe. Irish based companies trade approximately 15% of the infant milk formula tonnage internationally and it is anticipated that this will increase further to 20% in the coming years (Teagasc Annual Report, 2011) as global demand for infant and follow-on formula continues to rise in line with population growth. Currently, infant formula production contributes €620 million to the value of Irish dairy exports, which is greater than the contribution of either cheese (€600 million) or butter (€542 million) (Teagasc Publication, 2014).

Detailed milk product quality is not considered in the Irish national dairy cow-breeding objective, at present, despite its fundamental importance to potentially add value to the Irish agri-food industry. This is largely due to lack of routine access to data on detailed milk quality parameters, possibly owing to the expense of generating such data using currently available gold standard methods. Consideration of milk quality parameters in national breeding goals is particularly important for exporting countries such as Ireland to consistently achieve a high-quality product suitable for value-added international markets.

1.4 Milk protein composition

The high nutritional value of milk is partly due to its protein composition. Milk protein is a complex group of peptides in which over 200 different molecules have been characterised (Ng-Kwai-Hang, 2002). Bovine milk generally consists of about 3.3% protein, of which 78% is comprised of casein (CN), 17-18% of whey protein and the remaining 4-5% of non-protein nitrogen. Milk CN consists of alpha-s_1-casein (αs_1-CN), alpha-s_2-casein (αs_2-CN), beta-casein (β-CN), kappa-casein (κ-CN) (in approximate proportions of 4:1:4:1) and gamma-casein which is a product of degradation of β-CN (Ostersen et al. 1997; Miller et al. 1990). The three main whey proteins are beta lactoglobulin (β-LG), of α-lactalbumin (α-LA) and blood serum albumin, representing

approximately 50, 20 and 10% of total whey proteins, respectively. The remaining part of protein consists of non-protein nitrogen including immunoglobulins and trace amounts of several other proteins, including enzymes and growth factors (Fox and McSweeney, 1998).

Human and bovine milk have significantly different composition. Total whey to total casein protein ratio varies from 60:40 in human milk to 20:80 in bovine milk. Albeit infant formulas have been modified to have a ratio of total whey to total casein close to that expected in human milk, the concentration of α-LA is still relatively low in infant formula, whereas β-lactoglobulin A (β-LG A) (which is not present in human milk) is present in the greatest amount in bovine milk and is considered a major milk allergen (Jabed et al., 2012). In recent years, whey supplementation has allowed the development of infant formula with a greater concentration of α-LA and reduced concentration of β-LG A. Although whey-based formulas have a whey to casein protein ratio more similar to human milk than traditional milk formulas, the proportion of α-LA and β-LG A still substantially differ to the proportion in human milk (Lien, 2003).

Milk caseins are fundamental to the cheese making process as they form the gel network that captures the other constituents of cheese. Numerous studies have investigated the effects of milk protein polymorphisms, in particular those of CN on cheese yield (Wedholm et al., 2006; Bonfatti et al., 2011a). Wedholm et al. (2006) stated that the concentration of CN in milk protein had a favourable effect on the quantity of protein transferred from milk into cheese curd and high concentrations of α_{s1}-, β-, and κ-CN, and of β-LG B were found to significantly increase cheese yield.

Protein plays an important role in the immunity, growth and development of infants (Lönnerdal, 2003). Casein fractions supply essential amino acids and are carriers of phosphate and calcium, whereas, whey proteins are of nutritional benefit as they have a high concentration of essential amino acids such as methionine and cysteine (De Wit, 1998). Milk protein stimulates muscle synthesis, and some proteins and peptides in milk have positive health effects on blood pressure, inflammation, oxidation and tissue development (Haug et al., 2007). For example, the lactoferrin protein plays an important role in immune system maintenance due to its antibacterial, antifungal, and antiviral properties (Farnaud and Evans, 2003; Baker and Baker, 2005).

Table 1.1 Mean and standard deviation (SD) for protein composition from three different studies (two populations of Simmental cows and a population of Holstein-Friesian cows).

	Bonfatti et al. 2011a		De Marchi et al. 2009b		Rutten et al. 2011	
	n=1336 Simmentals		n=2167 Simmentals		n=5545 Holstein Friesians	
Trait	Mean	SD	Mean	SD	Mean	SD
Protein, g/L	40.68	4.50	40.12	0.12	35.10	0.30
Casein, g/L	35.61	3.95	35.10	0.12	75.19	1.73
Whey protein, g/L	5.06	0.71	-	-	-	-
Casein number, %	87.55	1.08	-	-	-	-
Casein, g/L	-	-	-	-	-	-
Alpha s1 casein	12.63	1.51	12.54	0.12	33.65	1.68
Alpha s2 casein	13.24	0.67	4.32	0.17	10.37	1.40
Beta casein	3.70	1.99	14.79	0.15	27.14	1.58
Gamma casein	4.37	0.42	-	-	-	-
Kappa casein	1.59	0.88	3.71	0.24	4.02	0.58
Glyco kappa casein	1.76	0.59	-	-	-	-
Whey protein, g/L	-	-	5.01	0.15	10.78	1.22
Alpha lactalbumin	1.29	0.22	1.30	0.18	2.44	0.32
Beta lactoglobulin	3.76	0.61	3.71	0.17	8.34	1.19

1.4.1 Milk protein determination

A range of methods exist to quantify milk protein composition including electrophoretic techniques (Ng-Kwai-Hang, 1984; Kroeker, 1985; Kim and Jimenez-Florez.,1994), isoelectric focusing (Kim and Jimenez-Florez., 1994), high performance liquid chromatography (HPLC) by ion exchange (Hollar et al., 1991), hydrophobic interactions (Bramanti et al., 2002), reversed-phase methods (Visser et al., 1991), and more recently capillary zone electrophoresis (Recio et al., 1997), mass spectrometry (Miralles et al., 2003) or these methods combined (Mollé and Léonil, 2005). Above all the reference methods, HPLC enables rapid and automated analysis, characterised by good separations, high resolutions and accurate, reproducible results.

1.5 Free amino acid composition

Free amino acids (FAA) in milk are amino acids (AA) resulting from milk protein denaturation and therefore do not contribute to the total protein of milk. Because FAA in milk processing arise from protein hydrolysis, it indicates poor quality milk.

1.5.1 Functions of amino acids

Amino acids play a role in muscle protein synthesis (valine; Ha et al., 2003), muscle growth (valine; Ha et al., 2003), antiviral activity (lysine; Habuka et al., 1989), protein methylation (lysine; Nakayama et al., 2001), transport of sulphur (cysteine; Stipanuk, 2004), haemoglobin structure and function (histidine; Lukin and Ho, 2004), and gene expression and immune function (glutamine; Curi et al., 2007); thus amino acids are often given as human nutritional supplements (Wu, 2009).

Human and bovine milk have different FAA content and composition, with bovine milk generally having a lesser concentration of FAA than human milk (Roucher et al., 2013; Agostoni et al., 2000; Sawar et al., 1997; Armstrong et al., 1963) and thus infant formula is supplemented with additional amino acids. Valine, histidine, and glutamic acid are essential in the production of infant formula (Koletzko *et al.,* 2005). The human body can produce glutamine from glutamic acid; however when under stress, body requirements for glutamine may surpass the ability of the body to produce sufficient amounts of glutamine. Consequently, some studies have suggested that glutamine may become a “conditionally essential” amino acid in the critically ill (Andrews and Griffiths 2002; Askanazi et al. 1980).

1.5.2 Free amino acid determination

The determination of FAA in milk can be difficult as FAA are normally only present in small quantities in the milk and are hard to detect in their natural forms. Common methods to quantify the FAA in bovine milk include;

1) Microchip Electrophoresis-Laser Induced Fluorescence Detection (MCE-LIF) (Wu et al., 2012). Advantages of this technique include that it is quick and inexpensive. A disadvantage of MCE-LIF that it has a low reproducibility.
2) Cation Exchange HPLC (Schwarz et al., 2005; Mounier et al., 2007). Advantages of Cation Exchange HPLC include that it enables rapid and automated analysis, characterised by good separations, high resolutions and accurate results. Disadvantages of Cation Exchange HPLC include low reproducibility and low sensitivity.

1.6 Milk colour

Raw milk colour influences the colour of subsequent dairy products and by-products (Descalzo et al., 2012) thereby influencing the attractiveness of the milk for different markets. A yellow colour of dairy products is considered unfavourable in Middle Eastern dairy markets (Keen and Wilson, 1992), but in Europe, a yellow colour is favourable in high fat dairy products such as butter and full fat cheeses (Hutchings 1994, Casalis et al., 1972). The yellow colour of bovine milk is related to the level of β-carotene and fat content in milk; a greater milk fat and β-carotene content results in an increase in the yellowness of milk (Tian et al., 2010) and of dairy products, because there is minimal loss of carotenoids when transforming milk to butter and cheese.

The white colour of milk is a function of the milk's physical structure; the dispersion of both casein micelles and fat globules in milk is responsible for the diffusion of incident light and is related to lightness (L*) (Raty et al., 1999). The natural pigmentations from carotenoids, protein and riboflavin are also associated with the white colour of milk. Milk with a low carotenoid, high protein and high riboflavin content tends to be whiter (Solah et al., 2007).

Feeding and selective breeding of cows may alter the carotenoid level and thus the colour of dairy products (Norieze et al., 2006a). Cows fed grass silage tend to produce milk with yellower fat and greater β-carotene content than milk produced by cows on a hay diet (Noziere et al., 2006b; Calderon et al., 2007). Breeds of cows such as Jerseys,

which produce milk with a greater carotenoid and fat content, produce more yellow colour milk than breeds such as Holstein-Friesians or Montbelliardes (Lucas et al., 2006). Differences in milk colour can also be related to the presence of abnormalities in milk; for example, mastitis attributable to *S. esculine* infection causes milk to have a more reddish/yellowish colour (Espada et al., 2002).

1.6.1 Milk colour determination

The two reference methods commonly used to determine colour include a visual colour system and a mathematical colour system. The visual colour system uses actual physical colour samples in a certain arrangement. An example of a physical colour system is the Munsell system, where milk is measured using a photometer or more recently spectrometers such as the Beckman (Nelson, 1948). Mathematical colour systems are used with a spectrophotometer and milk colour is related to precise descriptions of the light source, object and a standard observer (Welsh, 1993). Examples of mathematical colour systems are the Chroma Meter CR400 (Konica Minolta Sensing Europe, Edisonbaan 14-F, NE) or a NH310 Colour Meter Milk and Liquid Colour Test Machine (Shenzhen 3nh Technology Co., Ltd, China). Results from these systems are expressed according to the CIE Lab (L*a*b* colour space). The CIE Lab identifies colours by plotting the co-ordinates in a uniform colour space and by reporting values for lightness (L*), redness and greenness (a*), and yellowness and blueness (b*) (CIE, 1978). The L* coefficient ranges from black (L* = 0) to white (L*=100). The colour coefficients a* and b* represent true neutral gray values at a* = 0 and b* = 0. The colours red and green are represented along the a* axis, with green at negative a* values and red at positive a* values. The colours yellow and blue are represented along the b* axis, with blue at negative b* values and yellow at positive b* values.

1.7 Use of mid-infrared spectroscopy to phenotype milk

Traditionally used to predict total milk fat and protein content, MIRS has gained momentum as a potential tool to collect many more milk phenotypes in recent years. The number of peer-reviewed published manuscripts per year on the application of MIRS to predict milk related traits between 2006 to 2016 inclusive is provided in **Figure 1.3**. The rapid increase in manuscripts on the topic of MIRS emphasises the increasing interest in the topic. Such publications include studies on the use of MIRS to predict novel milk quality characteristics such as; milk fatty acid composition (Soyeurt et al., 2006, 2008,

2011; Rutten et al., 2009; De Marchi et al., 2011; Ferrand et al., 2011; Maurice-Van Eijndhoven et al., 2012), milk protein and CN percentage (Luginbuhl et al., 2002; Sorensen et al., 2003; Etzion et al., 2004) milk protein composition (De Marchi et al., 2009b; Bonfatti et al., 2011a; Rutten et al., 2011) milk coagulation properties (MCP; Dal Zotto et al., 2008; De Marchi et al., 2009a, 2013; Visentin et al., 2015), and milk acidity (De Marchi et al., 2009a). The usefulness of MIRS to predict animal-level characteristics such as energy balance (McParland et al., 2011, 2012), feed intake (McParland et al., 2011, 2012, 2014), feed efficiency (McParland et al., 2014) and methane emissions (Dehareng et al., 2012) have also been recently highlighted. The use of MIRS as a milk phenotyping tool has been the basis of several large-scale internationally collaborative research projects in recent times. Examples include Robustmilk (http://www.robustmilk.eu) and Optimir (http://www.optimir.eu).

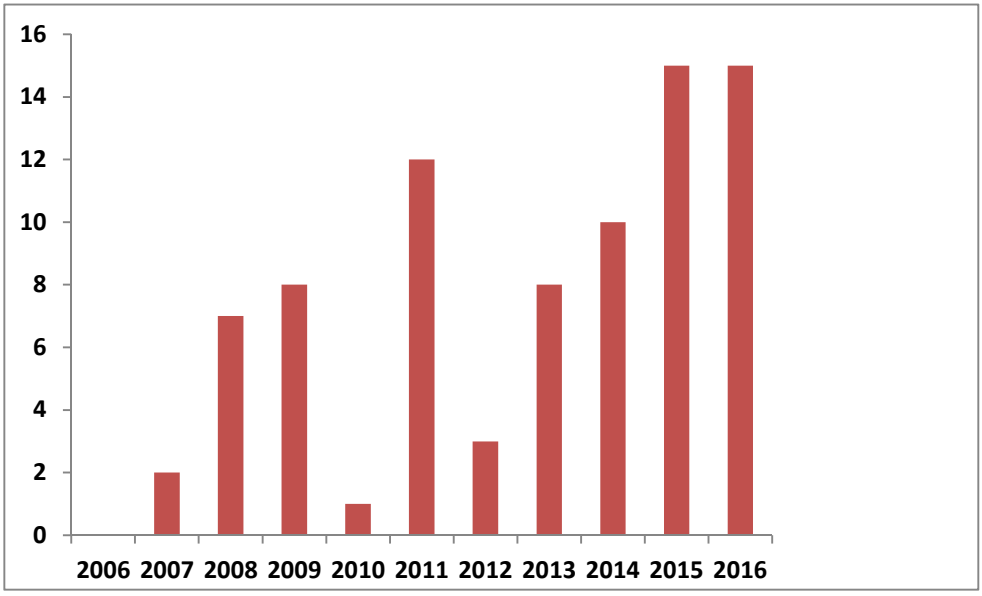


Figure 1.2 Published papers retrieved from ISI Web of Science on milk composition predicted by mid-infrared spectroscopy from 2006 to 2016.

1.7.1 Use of mid-infrared spectroscopy to predict protein composition

Studies that have investigated the use of MIRS to predict protein and CN composition of milk indicated excellent coefficient of determination (R^2) and root mean square error values of cross validation. Mean values for CN fractions varied across studies, for example mean values for α_{S1}-CN ranged from 12.63 g/L to 33.65 g/L (**Table 1.1**). Limited studies exist evaluating the effectiveness of MIRS in predicting milk protein fractions (Luginbuhl, 2002; Sorensen et al., 2003; Etzion et al., 2004; De Marchi et al., 2009b; Bonfatti et al., 2011a; Rutten et al., 2011). Prediction accuracies for CN fractions

ranged in value from an R^2 of 0.13 for β-CN (Bonfatti et al., 2011a; expressed as a percentage of protein) to an R^2 of 0.66 for α_{S1}-CN (De Marchi et al., 2009b; expressed as g/L of milk) (**Table 1.2**). The whey protein, α-LA, had the lowest R^2 value (0.20) (Rutten et al., 2011; using untreated spectra) and β-LG had the highest R^2 value (0.64) (Bonfatti et al., 2011a; using treated spectra). Regarding total whey protein and whey protein fractions, very similar accuracy of predictions were obtained by De Marchi et al. (2009b), Bonfatti et al. (2011a), and Rutten et al. (2011) (**Table 1.3**).

1.7.2 Factors influencing phenotyping of milk by MIRS

1.7.2.1 Reference method

On average, CN fractions were predicted more accurately from MIRS when the reference method was reversed-phase HPLC rather than capillary zone electrophoresis (De Marchi et al., 2009b; Bonfatti et al., 2011; Rutten et al., 2011), as HPLC is a more accurate reference method (Jimidar et al., 1993). The accuracy was greater using HPLC with milk samples having a lower standard deviation and achieving lower detection limits than capillary zone electrophoresis. In addition, the condition of the capillary can be difficult to keep constant between days of analysis in capillary zone electrophoresis.

1.7.2.2 Unit of measurement

Higher coefficient of determination values were obtained when protein fractions were expressed in absolute concentration (grams per litre of milk) rather than on a relative scale (percentage of protein or casein). Bonfatti et al. (2011a) obtained unsatisfactory results predicting CN fractions using relative values compared to traits expressed per unit of milk (**Table 1.2**), which verifies previous findings for MIRS prediction of fatty acid (FA) composition (Soyeurt et al., 2006). Similarly, MIRS prediction accuracies of whey fractions were greater when traits were expressed per g/L of milk (Bonfatti et al., 2011a).

1.7.2.3 Quantity of the gold standard and predict

Prediction accuracies were greater for total proteins compared to individual proteins, as total proteins were present in greater concentrations (De Marchi et al., 2009b). Soyeurt et al. (2006, 2011) and Rutten et al. (2009), who both attempted to predict milk FA content using MIRS, also reported greater accuracy of predictions for the components in greater concentration in milk.

Table 1.2 Unit of measurement and coefficient of determination (prediction error in parentheses) from cross validation of MIRS models to predict protein, casein, and casein fractions.

	Sorensen et al. (2003)	Etzion et al. (2004)	De Marchi et al. (2009b)	Bonfatti et al. (2011a)			Rutten et al. (2011)
Unit	%	%	g/l of milk	g/l of milk	% PRT	% CN	g/100g of milk
PRT[1]	-	0.94 (0.08)	0.58 (3.11)	0.78 (2.13)	-	-	-
CN[2]	0.97 (0.03)	-	0.58 (2.76)	0.77 (1.91)	-	-	0.25 (1.50)
αs_1-CN[3]	-	-	0.50 (1.07)	0.66 (0.89)	0.23 (1.95)	0.20 (2.34)	0.18 (1.52)
αs_2-CN[4]	-	-	0.35 (0.58)	0.49 (0.48)	0.17 (1.08)	0.19 (1.25)	0.26 (1.20)
β-CN[5]	-	-	0.33 (1.77)	0.53 (1.37)	0.13 (2.42)	0.16 (2.63)	0.19 (1.42)
κ-CN[6]	-	-	0.44 (0.68)	0.63 (0.55)	0.36 (1.44)	0.36 (1.62)	0.28 (0.49)

[1]PRT=Protein

[2]CN=Casein

[3]αs_1-CN = Alpha s_1 casein

[4]αs_2-CN = Alpha s_2 casein

[5]β-CN = Beta casein

[6]κ-CN = Kappa casein

Table 1.3 Unit of measurement, spectral pre-treatment applied[1], coefficient of determination, and prediction error (in parentheses) of cross validation procedures for MIRS prediction models of whey protein and whey fractions[2]

	De Marchi et al. (2009a)	Lopez-Villalobos et al. (2009)	Bonfatti et al. (2011)			Rutten et al. (2011)
Spectra Pre-treatment	Untreated	Untreated	SNV, De, MSC, 1D, 2D			Untreated
Unit	g/L of milk	mg/L of milk	g/L of milk	% PRT	% Whey PRT	g/100 g of milk
Whey protein	0.53 (0.51)	-	0.61 (0.45)	-	-	0.53 (0.84)
α-LA	0.29 (0.19)	-	0.31 (0.18)	0.31 (0.42)	-	0.20 (0.29)
β-Lg	0.55 (0.43)	-	0.64 (0.37)	0.42 (0.74)	0.36 (3.02)	0.56 (0.79)

[1]SNV = Standard normal variate, DE = Detrend, MSC = Multiplicative scatter correction, 1D = First order derivative, 2D = Second order derivative

[2]α -LA = Alpha lactalbumin, β-LG **=** Beta lactoglobulin A, PRT=Protein

1.7.2.4 Variability of reference sample

Milk samples from different breeds of cows, parities, stages of lactation, milking times, test dates and herds could reduce prediction accuracy as it increases the variability in the dataset, and traits are then harder to predict by partial least square regression. However, it will make the equations more applicable and not just applicable to the dataset that the equations were developed on. Therefore, the ultimate aim using MIRS is to develop prediction equations on a dataset with a large amount of variation but still obtain high prediction accuracies, thus the prediction equations can be used in dairy industry.

1.8 Analysis of mid-infrared spectroscopy data

1.8.1 Mathematical pre-treatments

Derivatives are among the most common signal pre-treatments applied to spectral data. Principally, derivatives are used to resolve peak overlap (or enhance resolution) and to bring all spectra to a common baseline. The use of spectral mathematical pre-treatments such as the first and second derivative in IR spectroscopy, make it more feasible to determine chemical composition and correlated compounds (**Figure 1.4).** Use of higher derivatives is not advantageous for calibration model development because as the derivative increases from one to four, so will the associated noise; therefore, the first and second derivatives are the most frequently used. Spectral derivatives are calculated by obtaining the differences between two consecutive absorbance points, or between pre-defined gap distances. The first derivative is simply the slope of the absorbance spectra and uses two wavelengths points. The following equation (**Eq.1**) shows how the first derivative is calculated (Williams, 2007). The second derivative uses three wavelength points. Twice the values of the data at point B are subtracted from the data at point A, and the data from the point C are added to the result. The following equation (**Eq. 2**) demonstrates how it is calculated.

Eq.1 1^{st} Derivative = Absorbance of wavelength 1 - Absorbance of wavelength 2

Eq.2 2nd Derivative = Absorbance of wavelength 3 – [(2 x Absorbance of wavelength 2) + Absorbance of wavelength 1]

First derivatives remove the effect of baseline shifts caused by variable packing density and particle size. Models based on the first derivative may give better results than the second derivative as they are developed from the sides of absorption bands, but the

spectra can be difficult to interpret. Noise is defined as any electronic signal that reaches the detector that is not directly related to the actual absorption bands required for calibration and subsequent analysis. Second derivatives orientate bands downwards and are helpful in detecting what may be causing the main absorption, as it sharpens the absorption bands, which makes it easier to identify bands that have overlapping absorptions. However, the amount of noise increases as the derivative increases leading to poorer prediction accuracies (Williams, 2007).

1.8.2 Principal component analysis

Principal component analysis (PCA) is a statistical multivariate analysis technique, which captures the correlation among variables and represents data as a new set of fewer variables explaining the maximum variance. These variables are denoted as principal components (PCs) and each PC is a linear combination of the original variables (Jolliffe, 2002). The aim of principal component analysis is to explain the maximum amount of variance with the fewest number of PCs, for easy exploration and further analysis, such as regression, clustering, and discriminant analysis. It is concerned with explaining the variance-covariance structure of a set of variables through a few new variables. All principal components have three important properties (Amnarttrakul and Thongteeraparp, 2011), which are:

1) The PCs are uncorrelated.

2) The first PC explains the greatest variance; the second PC explains the second greatest variance, etc.

3) The total variation of all PCs combined is equal to the total variation of the original variables.

Observations that are outliers with respect to the first few PCS or the major PCs usually correspond to outliers on one or more of the original variables. Outliers in the original variables can be detected by making a bivariate scatterplot of the first and second PCs.

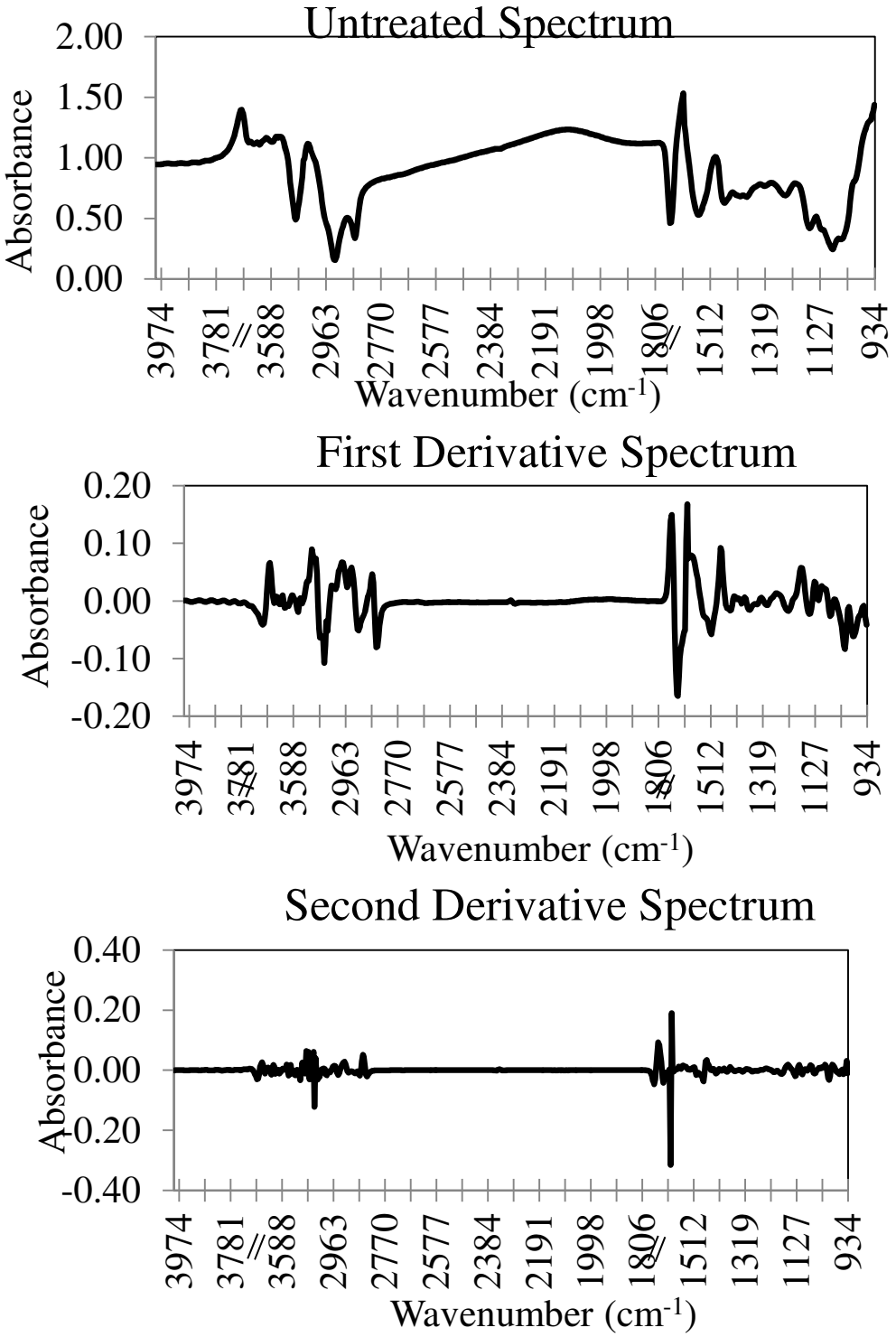


Figure 1.3 Example of untreated, first derivative and second derivate spectra (De March et al., 2014).

1.8.3 Partial least square regression

Partial least squares (PLS) regression is a recent technique that generalises and combines features from PC analysis and multiple regression (Helland, 1990). It is particularly useful to predict a set of dependent variables from a large set of independent

variables (i.e., predictors). The PLS regression is a method of constructing predictive models when they are many factors that are highly collinear. It can be a useful tool when prediction is the goal of analysis and there is no need to limit the number of measured factors.

The goal of PLS regression is to predict Y from X and to describe their common structure. When the number of predictors is large compared to the number of observations, X is likely to be singular and the ordinary multiple regression approach is no longer feasible (i.e., because of multicollinearity). PLS regression finds components from X that are also relevant for Y. Precisely, PLS regression looks for a set of components that performs a simultaneous breakdown of X and Y with the constraint that these components explain the maximum amount of the covariance between X and Y. This is followed by a regression step where the breakdown of X is used to predict Y (Abdi et al., 2003). Some of the software programmes commonly used to handle MIRS data include: SAS (SAS Institute Inc., Cary, NC), Unscrambler (Camo, Norway), simca-p (mks, Sweden), Winisi (Foss Electronic A/S, Hillerød, Denmark) and PLS toolbox (Eigenvector Research, Inc., Manson).

1.9 Factors associated with protein fractions and free amino acids

The ratio of protein fractions in milk affect various processing attributes of the milk including, rennet coagulating time (Ikonen et al., 2004; Joudo et al., 2008), curd firmness (Ikonen et al., 2004; Wedholm et al., 2006, Joudo et al., 2008), pH (Ikonen et al., 2004; Joudo et al., 2008) and cheese yield (Wedholm et al., 2006; Bonfatti et al., 2011a). Both genetic and management factors influence the quantity of individual milk proteins and FAA of bovine milk. Very few studies have investigated factors associated with FAA in milk.

1.9.1 Animal characteristics influencing milk quality

1.9.1.1 Parity

Younger animals had a greater concentration of αS-CN compared to their older contemporaries and β-CN decreased as parity number increased (Kroeker et al., 1985). Younger animals also had a greater concentration of the whey fraction β-LG compared to

their older contemporaries (Ng-Kwai-Hang et al., 1987). To our knowledge, no previous study exists which investigated the association of parity with FAA or provided estimates of heterosis for FAA.

1.9.1.2 Stage of lactation

Stage of lactation also influenced the content of individual protein fractions in milk (Ng-Kwai-Hang et al., 1987; Kroeker et al., 1984; Ostersen et al., 1977). The protein fractions of αS-CN, β-CN and κ-CN, β-LG decreased in early lactation followed by an increase throughout the remainder of lactation. Changes in proportions of αS-CN, β-CN and κ-CN, β-CN, β-LG and α-LA in total milk protein content according to stage of lactation have also been demonstrated (Kroeker et al., 1985). The concentration of total FAA has been shown to be greatest in early and late lactation (Auldist et al., 1995), when milk bovine quality is poorest (Davis et al., 1994).

1.9.1.3 Breed

Significant breed differences existed for milk protein, CN and protein fractions (Cerbulis et al., 1975; McLean et al., 1984; Auldist et al., 2004; Joudu et al., 2008; Lopez-Villalobos 2012). Breeds analysed included Jersey, Holstein Friesian, Brown Swiss, Guernsey, Ayrshire, and Milking Shorthorn. Regarding protein percentage, breeds ranked from highest to lowest: Jersey (4.07%), Brown Swiss (3.84%), Guernsey (3.56%), Ayrshire (3.30%), Milking Shorthorn (3.17%), and Holstein (3.07%) (Cerbulis et al., 1975). Breeds differed in all other components and in milk yield with Brown Swiss ranked highest in yield of milk (6064kg), protein (233kg per lactation) and CN (191kg per lactation). Jersey breed had lower milk yields compared to Holstein Friesian breed (15.22 kg vs 21.71kg) but higher protein fraction contents (14.32g/kg vs 12.87g/kg for β-CN; Mc Lean et al., 1984).

1.9.1.4 Heterosis

Heterosis is defined as the increased performance of crossbred animals compared with the average of both purebred parental breeds (Sorensen et al., 2008). Previous studies have also demonstrated that heterosis had a significant positive effect on total protein content (Dechow et al., 2007; Bryant et al., 2007). A study by Back and Lopez–

Villalobos (2007) showed that heterosis had a negative effect on κ-CN and a positive effect on α-LA (P<0.05).

1.9.1.5 Recombination

Recombination loss is defined as the disintegration of epistatic effects to form nonparent inter-loci combinations of alleles in crossbred animals (Cassady et al., 2002). A study by Dechow et al. (2007) demonstrated that recombination had an unfavourable effect on protein (-3.31%).

1.10 Genetic parameters of milk quality

1.10.1 Heritability estimates

Heritability is the proportion of variance in a trait in a population that is attributable to genetic variation (Wray and Visscher, 2008) and is calculated as the ratio of additive genetic variance to phenotypic variance. Heritability estimates for milk protein fractions have increased in recent years and results differed across studies depending on the gold standard method, breeds, and sample size used (Schopen et al., 2009). Heritability estimates for total protein, CN, and CN fractions are shown in **Table 1.5**. In general, heritability estimates for protein fractions expressed as a percentage of total protein were greater than those for protein fraction contents expressed as a total amount in the milk (Bonfatti et al., 2011a).

Heritability has been estimated for milk production traits across DIM and parities, showing that fat and protein contents were more heritable in mid to late lactation, and more heritable in first parities compared to later parities (Bastin et al., 2011; 2012). To our knowledge, there is no study that estimates heritability for proteins predicted by MIRS, or for FAA with either gold standard determination or MIRS prediction.

Table 1.4 Differences in concentration of proteins in milk from Friesian and Jersey cows obtained in various studies (Lopez-Villalobos 2012).

	Auldist et al. (2004)		Back and Lopez-Villalobos (2007)		McLean et al. (1984)	
Trait[1]	Friesian (N=29)	Jersey (N=29)	Friesian (N=20)	Jersey (N=20)	Friesian (N=238)	Jersey (N=262)
Milk yield (kg)			21.71	15.22**	31.77	31.68
Protein (g/kg)	35.50	39.80**	35.70	40.60**	31.60	39.90**
Casein (g/kg)	27.40	31.20**	27.50	32.20**	24.50	31.20**
Casein: Protein	0.771	0.785**	0.77	0.793**	0.775	0.782
αs_1-CN (g/kg)	8.50	10.00*				
as_2-CN (g/kg)	2.80	4.00				
a-CN (g/kg)	11.50	11.90	16.67	19.97**		
β-CN (g/kg)	11.00	13.50**	12.87	14.32**	8.90	10.80*
κ-CN (g/kg)	3.80	4.10*	4.17	5.55**	2.60	3.90**
γ-CN (g/kg)	1.90	2.50				
α-La (g/kg)	1.30	1.50	0.90	1.02*	0.95	1.13
β-Lg (g/kg)	4.90	5.30*	4.64	5.03**	2.96	3.67

[1] αs_1-CN = Alpha s_1 casein, αs2-CN = Alpha s_2 casein, a-CN = Alpha s_1 casein + alpha s_2 casein, β-CN = Beta casein, κ-CN = Kappa casein, γ-CN = Gamma casein, α-la = Alpha lactalbumin, β-LG = Beta lactoglobulin (A + B)

[2] * = P< 0.05 concentration of proteins in milk from Friesian and Jersey within a study significantly different to each other

** = P<0.01 concentration of proteins in milk from Friesian and Jersey within a study significantly different to each other

Table 1.5 Estimates of heritability for milk protein fractions (Lopez-Villalobos 2012)

Trait	Kreoker et al., 1985[1]	Bobe et al., 1999[1]	Graml and Pirchner 2003[1]	Schopen et al., 2009[1]	Bonfatti et al., 2011b[1]	Huang et al., 2012[3]
Kappa casein	0.01	0.28	0.22-0.28	0.64	0.63	0.66
Alpha s_2 casein	-	0.01	0.17	0.73	0.28	-
Alpha s_1 casein	0.02	0.18	0.27-0.37	0.47	0.68	0.33
Beta casein	0.03	0.01	0.32-0.34	0.25	0.69	0.33
Alpha lactalbumin	0.14	0.00	0.22-0.26	0.55	-	0.33
Beta lactoglobulin	0.24	0.36	0.26-0.35	0.80	0.34	0.69
Casein[2]	-	-	-	0.41	-	-
Whey[3]	-	-	-	0.71	-	-
Casein index[4]	-	-	-	0.70	-	-

[1] Protein fractions were expressed as a percentage of the total protein

[2]Casein = alpha s_1 casein + alpha s_2 casein + beta casein + kappa casein; Whey = alpha lactalbumin + beta lactoglobulin; Casein index = casein/ (casein + whey) x 100

[3]Proteins fractions expressed as mg/100µL of milk

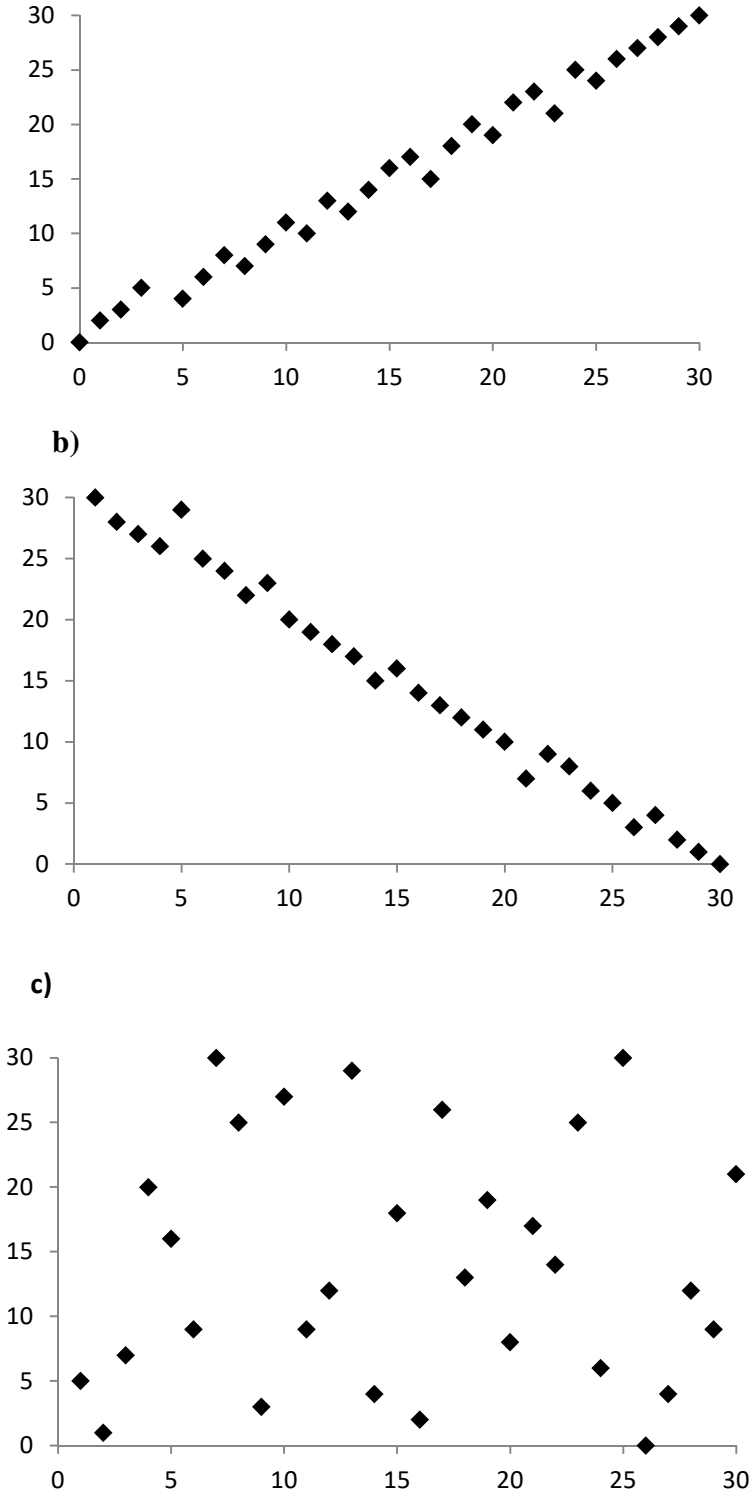


Figure 1.4 The strength of the correlation between two traits; a) a strong positive correlation, b) a strong negative correlation and c) no correlation exists.

1.10.2 Genetic correlations between protein fractions and between free amino acids

A correlation measures the strength of the linear relationship between two variables. Correlations are unit-less and constrained to between -1 and +1 **(Figure 1.4).** A positive correlation indicates that as one trait increases the second trait also increases and a negative correlation indicates that as one trait increases the second trait decreases (Van Vleck et al., 1987). A genetic correlation is a correlation between an animal's breeding value for one trait and the same animal's breeding value for another trait (Searle, 1961).

Due to the gold standard method for protein composition determination being complicated, laborious, and expensive, only 9 studies to date have attempted to quantify genetic variation in milk proteins (summarised by Bobe et al., 1999; Graml and Pirchner, 2003; Schopen et al., 2009). Only five of these studies reported estimates of genetic correlations for milk protein composition (Renner and Kosmack, 1975; Schopen et al., 2009). Schopen et al. (2009) found that most of the genetic correlations among major milk proteins in Dutch Holstein Friesians were weak and that β-LG concentration was strongly negatively correlated with the proportion of CN in milk; CN is imperative for cheese production. Bonfatti et al. (2011b) reported low genetic correlations among CN fractions, and between CN and whey protein fractions in Simmental cattle. Genetic correlations between five CN fractions (expressed in g/L) ranged from -0.14 (γ-CN and as_2-CN) to 0.56 (γ-CN and B-CN), however, when fractions were expressed as a percentage of total CN, genetic correlations ranged from -0.68 (αs_1-CN% and β-CN%) to 0.38 (αs_2-CN% and γ-CN%) (Bonfatti et al., 2011b). These results disagreed with those obtained by Schopen et al. (2009) who reported that the strongest genetic correlations were the ones between αs_1-CN% and αs_2-CN% and between αs_1-CN% and κ-CN%. However, actual differences between phenotypes in the studies were not large for example, Bonfatti et al. (2011b) reported an average value of 35.61% (SD, 2.65) for αs_1-CN% and Schopen et al. (2009) obtained a similar mean of 33.62% (SD, 1.70).

Findings by Bonfatti et al. (2011b) showed that selecting for increased total CN, increased the relative proportion of β-CN and decreased the proportion of αs_1-CN and αs_2-CN in milk. The estimated genetic correlations among milk protein fractions noticeably increased and many became significant when the statistical model accounted for protein gene effects. The genetic correlations among protein fractions αs_1-CN, β-CN, and κ-CN changed from a low or slightly negative correlation (-0.14) to a large and

positive correlation (0.95). This suggested that synthesis of all protein fractions undergo a shared regulation and that weak genetic relationships may arise from a differential transcriptional regulation (Bevilacqua et al., 2006).

In the literature, estimates for genetic relationships between milk coagulating properties, protein and CN percentage vary. Lindstrom et al. (1984) demonstrated that short coagulation time correlated with high protein percentage; however, in other studies short coagulation time correlated with low protein and CN percentage (Oloffs et al., 1992; Ikonen et al., 1999), or no correlation existed, between coagulation time and protein percentage (Oloffs et al., 1992, for Angler cows). According to Oloffs et al. (1992), high values for curd firmness correlated with high protein and CN percentage, whereas Ikonen et al. (1999) demonstrated that high values for curd firmness correlated with low protein and casein percentage. In a later study by Ikonen et al. (2004), the correlations between curd firmness and protein percentage and between curd firmness and CN percentage were negligible.

Results from Bonfatti et al. (2011b) demonstrated that a low milk pH correlated with favourable MCP, consistent with the findings of Ikonen et al. (2004). Rennet coagulating time was positively genetically correlated with αs_1- and αs_2-CN in CN, but was negatively correlated with the proportion of β-CN in CN (Bonfatti et al., 2011b). Weak curds were genetically related to increased proportions of αs_1-CN and αs_2-CN and decreased proportions of κ-CN in CN (Bonfatti et al., 2011b). Therefore, the use of selective breeding to increase the level of β-CN and κ-CN in milk, while decreasing the level of αs_1-CN and αs_2-CN in milk and milk pH, could have beneficial effects on MCP.

1.11 Gaps in knowledge

Gaps in knowledge that will be examined as part of this thesis include:

1) Effectiveness of MIRS in predicting milk protein fractions, FAA and colour in milk.
2) Cow level and other factors associated with protein fractions, FAA and colour in milk from grazing dairy cows.
3) Genetic parameters for protein fractions, FAA and colour in milk from grazing dairy cows.

CHAPTER 2

Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristic

A. McDermott*†, G. Visentin *† , M. De Marchi†, D.P. Berry*, M.A. Fenelon‡, P.M. O'Connor‡, O.A. Kenny‡ and S. McParland*

*Teagasc Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

†Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16,35020 Legnaro (PD), Italy

J. Dairy Sci. 99:3171–3182

2.1 Abstract

The aim of this study was to evaluate the effectiveness of MIRS in predicting milk protein and free amino acid composition in bovine milk. Milk samples were collected from seven Irish research herds and represented cows from a range of breeds, parities, and stages of lactation. Mid-infrared spectral data in the range of 900-5,000 cm^{-1} were available for 715 milk samples; gold standard methods were used to quantify individual protein fractions and FAA of these samples with a view to predicting these gold standard protein fractions and FAA levels with available mid-infrared spectroscopy data. Separate prediction equations were developed for each trait using partial least squares regression; accuracy of prediction was assessed using both cross validation on a calibration data (n=400 to 591 samples) and external validation on an independent data set (n=143 to 294 samples). The accuracy of prediction in external validation was the same irrespective of whether undertaken on the entire external validation dataset or just within the Holstein-Friesian breed. The strongest coefficient of correlation obtained for protein fractions in external validation was 0.74, 0.69, and 0.67 for total casein, total beta lactoglobulin, and beta casein, respectively. Total proteins (i.e., total casein, total whey and total lactoglobulin) were predicted with greater accuracy then their respective component traits; prediction accuracy using the infrared spectrum was superior to prediction using just milk protein concentration. Weak to moderate prediction accuracies were observed for FAA. The greatest coefficient of correlation in both cross validation and external validation was for Gly (0.75), indicating a moderate accuracy of prediction. Overall, the FAA prediction models over-predicted the gold standard values. Near unity correlations existed between total casein and beta-casein irrespective of whether the traits were based on the gold standard (0.92) or mid-infrared spectroscopy predictions (0.95). Weaker correlations among FAA were observed than the correlations among the protein fractions. Pearson correlations between gold standard protein fractions and the milk processing characteristics of rennet coagulation time, curd firming time, curd firmness, heat coagulating time, pH and casein micelle size were weak to moderate and ranged from -0.48 (protein and pH) to 0.50 (total casein and a30). Pearson correlations between gold standard FAA and these milk processing characteristics were also weak to moderate and ranged from -0.60 (Valine and pH) to 0.49 (Valine and K_{20}). Results from this study indicate that mid-infrared spectroscopy has the potential to predict protein fractions and some FAA in milk at a population level.

2.2 Introduction

Detailed milk product quality is not considered in the Irish national dairy cow breeding objective, at present, despite its fundamental importance for adding value to the Irish Agri-food industry. This is simply due to lack of routine access to data on detailed milk quality parameters, possibly owing to the expense of generating such data using gold standard methods. Consideration of milk quality parameters in national breeding goals is particularly important for exporting countries such as Ireland to consistently achieve a high quality product, suitable for value added international markets.

The concentration of protein and the composition of protein fractions in milk influence the production efficiency of cheese, infant milk formula, and both casein and whey protein supplements. Wedholm et al. (2006) stated that the concentration of casein in milk protein has a favourable effect on the quantity of protein transferred from milk into cheese curd and high concentrations of α_{S1}-, β-, and κ-CN and of β-LG B were found to significantly increase cheese yield. Elofsson et al. (1996) demonstrated a low β-Lg concentration reduces the fouling rate of heating equipment. Beta lactoglobulin, which is not present in human milk, is a major milk allergen and therefore efforts have been made to reduce the level of this protein in cow milk (Jabed et al., 2012). Therefore, milk protein composition is of increasing importance to the dairy industry due to the expected global demand for cheese (FAO, 2014). Protein plays an important role in immunity, growth and development of infants (Lönnerdal, 2003). Therefore, milk protein composition is particularly important for infant formula production (De Wit, 1998) as the composition of bovine milk is different to human milk (Jensen, 1995). Infant formula production is the fastest growing sector in the world dairy market (FAO 2014) and the international market for infant milk formula is worth approximately US$5-6 billion annually. Protein composition also affects milk processing characteristics such as the heat coagulating time of bovine milk (Singh, 2004).

Regarding milk processing ability, high free amino acid levels indicate poor quality milk as they arise from protein hydrolysis and are generally in greatest concentration in early and late lactation milk (Davis et al., 1994), when milk quality is poorest (Auldist et al., 1995). Human and bovine milk have different FAA content and composition, with bovine milk generally having a lesser concentration of FAA than human milk (Armstrong et al., 1963; Sawar et al., 1997; Agostoni et al., 2000; Roucher et al., 2013). Therefore for nutritional reasons, supplementation of infant formula with the

required FAA may be of interest in infant formula production. Achieving a milk FAA profile in bovine milk, similar to that of human milk through breeding may be an alternative strategy. The FAA profile of milk is therefore of interest to dairy farmers, as milk processors may pay higher prices for milk based on its FAA composition.

Milk compositional traits such as protein fractions have a major influence on milk processing ability-related traits such as rennet coagulating time (Auldist et al., 2004; Ikonen et al., 2004; Wedholm et al., 2006). It is well documented that milk composition and milk coagulation properties are affected by environmental factors including stage of lactation (Heck et al., 2009; Ostersen et al., 1997). Auldist et al. (1995) documented a stage of lactation effect on both cheese yield and quality. The majority of milk production in Ireland is seasonal (Berry et al., 2006), as most dairy cows calve in spring (Berry et al., 2013). Therefore, it may be of interest to milk processors on how the correlations between milk compositional traits and processing ability characteristics differ in different stages of lactation.

Despite the importance of quantifying individual proteins and FAA in milk, no inexpensive and efficient method of measuring these components in milk is available. Mid–infrared spectroscopy (MIRS) is a technique that studies the interactions between light and matter at wavelengths in the spectral range of 900 to 5000cm^{-1}. It is based on the capability of molecules to reflect, transmit or absorb part of the electromagnetic radiation when exposed to light. According to the Beer-Lambert law (Swinehart, 1962), the quantity of the electromagnetic radiation absorbed is directly proportional to the amount of the absorbent molecule in the sample. Mid–infrared spectroscopy is an efficient method currently used by milk recording organizations worldwide to predict milk fat, protein, and lactose and has recently been used to predict more detailed milk composition traits such as fatty acids (Soyeurt et al., 2011; De Marchi et al., 2011), coagulation traits (De Marchi et al., 2013) as well as animal-level characteristics such as energy balance (McParland et al., 2011, 2012) and feed efficiency (McParland et al., 2014). Limited studies exist evaluating the effectiveness of MIRS in predicting milk protein fractions (Bonfatti et al., 2011a; Rutten et al., 2011; De Marchi et al., 2010). The gold standard method used in both the studies of Bonfatti et al. (2011a) and De Marchi et al. (2010) was high performance liquid chromatography (HPLC); however Rutten et al. (2011) used capillary zone electrophoresis. In the studies of Bonfatti et al. (2011a) and Rutten et al. (2011), the ratio performance deviation ranged from 1.04 (gamma casein) to

2.12 (protein) and from 0.48 (beta casein) to 1.06 (total whey), respectively. Across studies (Bonfatti et al. 2011a; Rutten et al. 2011), the coefficient of determination for cross validation ranged from 0.08 (gamma casein) to 0.80 (protein). However, De Marchi et al. (2009) expressed protein fractions as grams per litre, whereas Rutten et al. (2011) expressed them on a protein percentage basis (g/100g) and Bonfatti et al. (2011a) expressed them in both forms. Higher coefficient of determination values were obtained when protein fractions were expressed in grams per litre rather than on a percentage basis. The aim of this study was to quantify the effectiveness of MIRS to predict individual milk proteins and FAA as well as to estimate the association between these MIRS-predicted traits and other phenotypic characteristics of milk including rennet coagulating time (RCT), curd firming time (k_{20}), curd firmness (a_{30}) heat coagulation Time (HCT) and pH. The use of MIRS as a tool to predict detailed milk quality traits is attractive since the MIR spectrum of individual milk samples is available at a negligible cost to routine milk recording.

2.3 Materials and methods

2.3.1 Milk sample collection

Milk samples were obtained from seven research farms operated by the Teagasc Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork between August 2013 and August 2014, inclusive. Cows were milked daily at 07:00 h (AM) and 15:00 h (PM) and milk composition was recorded weekly using a MilkoScan FT6000 (Foss Electronic A/S, Hillerød, Denmark). The resulting spectrum, containing 1,060 transmittance data in the mid-infrared region between 900 and 5,000 cm^{-1}, was stored. Following composition analysis, 505 AM and 225 PM milk samples were preserved with Broad Spectrum Microtabs II containing 8 mg of Bronopol and 0.3 mg of Natamycin (D&F Control Systems Inc., Norwood, MA, USA) and stored at 4 °C for further analysis. Samples were selected to maximise diversity of breed [Holstein Friesian (n=454), Jersey (n=117), Norwegian Red (n=15) and Holstein Freisian, Jersey and Norweigan Red crossbreds (n=144)], stage of lactation, milking time (i.e AM or PM milking), parity and MIR spectrum and represented 621 animals; animals had a maximum of three records each.

2.3.2 Gold standard methodologies

2.3.2.1 Milk protein determination

Total protein was predicted by MIRS and calibrated using the Kjeldahl method. Milk protein fractions were also determined for 557 samples within 48 hours of sample collection. Milk protein fractions were quantified using reversed phase High Performance Liquid Chromatography (RP-HPLC). Samples including the Sigma Aldrich standards were prepared in denaturing buffer (7 M urea + 20 mM Bis-tris propane, pH 7.5) to which 5 ul⁄mL Mercaptoethanol was added to give a final protein concentration of approx. 2 mg/ml. The samples were then incubated for 1 hour at room temperature before filtering through a 0.22 filter. Protein composition was determined by reverse-phase high-performance liquid chromatography using an adaptation of the method of Visser et al. (1991). Separation was performed using an Agilent Poroshell 300SB C18 column (2.1 mm × 75 mm; Agilent Technologies UK Ltd.). The HPLC system consisted of an Agilent 1200 Separation Module with MWD Detector and Agilent Chemstation Software. Gradient elution and peak detection were performed according to Reid et al. (2015) and Mounsey and O'Kennedy (2009). All casein and whey standards were supplied by Sigma-Aldrich.

Total CN was calculated as the sum of alpha $_{S2}$ casein (α_{S2}-CN), alpha $_{S1}$ casein (α_{S1}-CN), beta casein (β-CN) and kappa casein (k-CN); total whey was calculated as the sum of alpha lactalbumin (α-La), beta lactoglobulin A (β-Lg-A), beta lactoglobulin B (β-Lg-B).

Protein fractions were expressed as g/L of milk, but were also expressed as a percentage of total protein or casein, by dividing the yield of each protein fraction by the total protein or casein content of the milk sample.

2.3.2.2 Free amino acid determination

The FAA, Lys, Val, Glu, Gly, Asp, Arg and Ser were quantified in 715 milk samples using Cation exchange HPLC coupled with post column ninhydrin detection as described by Mounier et al. (2007). Seven hundred and fifty microliters of each milk sample was deproteinised by mixing with 750 µl of 24% (w/v) tri-chloroacetic acid and left to stand for 10 mins. Samples were subsequently centrifuged at 20,817x g (Microcentaur; MSE, London, United Kingdom) for 10 minutes (4 degrees Celsius). The resulting supernatants were diluted with 0.2 M sodium citrate buffer (pH 2.2) to give

approximately 250 nmol/ml of each amino acid residue. Samples were then diluted one in two with the internal standard, norleucine, to give an end concentration of 125 nm/ml. Twenty microliters of each sample was then quantified for FAA using a JEOL JLC-500/V amino acid analyzer (JEOL UK Ltd., Garden City, Herts., United Kingdom) fitted with a JEOL Na+ high-performance cation-exchange column.

2.3.2.3 Determination of milk coagulation properties

Milk coagulation properties were determined on preserved milk samples within five days of collection, using a Formagraph (Foss Electronic A/S, Hillerød, Denmark) as described by Visentin et al. (2015). Coagulation properties measured included (i) RCT, defined as the number of minutes taken from rennet addition to the beginning of the coagulation, (ii) k_{20}, the time from the gel development to a width of 20 mm in the graph, and (iii) curd firmness measured as the width of the graph after 30 minutes (a_{30}) after rennet addition.

2.3.2.4 Heat coagulation time and pH determination

Heat coagulation time (HCT) was tested within 48 hrs of sample collection using the hot oil bath method as described by Davies and White. (1966). Heat coagulating time was measured by visual analysis and taken as the time when each sample started to coagulate. Samples with a HCT >30 min were classified as non-coagulating and discarded from the analyses. The pH of each sample was measured using a Seven compact pH-meter S220 (Mettler Toledo AG, Switzerland) within 24 hrs of samples collection.

2.3.2. Data analysis

Identification of outlier gold standard values and trait distribution was determined using PROC UNIVARIATE in SAS (SAS Institute Inc., Cary, NC, USA). Traits that did not have a normal distribution were transformed using a natural logarithm transformation. Gold standard values that were > 3 standard deviations from the mean were considered to be outliers and up to three outliers were removed from the protein fraction analyses while up to 22 outliers were removed from the FAA analyses. Observations for each protein or FAA did not exist for all samples, primarily for logistical reasons (**Table 2.1**).

Spectral data were transformed from transmittance to linear absorbance through a logarithmic transformation of the reciprocal of the wavelength values (Soyeurt et al.,

2011). Preliminary analyses revealed no improvement in model prediction accuracy following mathematical pre-treatment (Savitzky-Golay first and second derivatives of the log transformed spectral data); therefore the prediction models were developed using untreated spectra. Only one spectrometer was used in the present study. Equations were developed to predict each milk quality parameter separately using partial least squares regression (Proc PLS; SAS Institute Inc., Cary, NC). Spectral regions from 926-1,580 cm^{-1}, 1,717-2,986 cm^{-1} and 3,696-3,808 cm^{-1}, were used to develop all prediction models based on the observed loadings for each wavelength.

Accuracy of the prediction equations was determined using external validation whereby 25% of data were excluded from equation calibration and used as an independent validation data set (VD). This procedure was repeated 4 times, using a different 25% of the data in the VD each time. Samples were selected for VD to represent similar variation to that present in the gold standard data in the calibration dataset (CD) used to develop prediction equations. For each prediction model, the dataset was sorted by the trait of interest. The first sample and every fourth sample thereafter were included in the VD for the first iteration; for the second iteration, the second sample and every fourth sample thereafter was chosen for the VD with a similar procedure used for the third and fourth iteration.

Therefore, separate VD and CD were generated for each prediction equation. All records from cows included in the VD were removed from the CD and included only in the VD; therefore no cow was represented in both the CD and VD in a given iteration.

Criteria used to determine the effectiveness of MIRS predictive models were the coefficient of correlation of cross validation (r_c) and external validation (r_v), the root mean square error of cross validation (RMSEc) and external validation (RMSEv), the slope (b), which is the linear regression coefficient between gold standard and MIRS-predicted values of each trait, the mean bias of prediction, which is the average difference between MIRS-predicted values and gold standard values in external validation, the standard error (SE) of the slope and the bias and the ratio performance deviation (RPD), which is the ratio of the standard error in prediction to the standard deviation of each trait. Four validation datasets were created and then appended onto each other and the r_c, r_v, RMSE, b (SE), and bias (SE) was calculated based on all four iterations of combined. The average number of factors (#L) used to build the prediction equations, was the average number from all four iterations rounded to the nearest whole number. Validation was also

performed within just the Holstein-Friesian breed (i.e., the predominant breed) as well as across breeds. Furthermore, protein fractions were also predicted in the external validation using just the total milk protein concentration and compared to prediction accuracy using the MIRS.

Pearson correlations among the gold standard and among the MIRS-predicted values of protein fractions and FAA were estimated. Pearson correlations between gold standard and MIRS-predicted FAA with RCT, k_{20}, $a_{30,}$ HCT and pH were also estimated.

2.4 Results

The total data set comprised of 730 samples; 584 milk samples were from spring-calving cows fed a predominantly grazed grass based diet and the remaining 146 samples were from autumn-calving cows fed a total mixed ration diet. Milk samples represented different stages of lactation and ranged from 5 to 375 days in milk; first to eleventh parity cows were represented.

Table 2.1 Number of records (n), mean, standard deviation (SD) and coefficient of variation (CV) for the studied traits.

Trait	n	Mean	SD	CV
Protein g/L				
Total CN	554	35.97	7.11	19.77
α_{S1}-CN	557	13.92	3.18	22.84
α_{S2} CN	555	3.62	0.97	26.90
β-CN	555	12.64	2.64	20.91
κ-CN	556	5.92	1.67	28.27
Total Whey	549	6.08	1.79	29.45
A-LA	551	1.11	0.32	28.34
Total Lg	552	4.97	1.65	33.20
B-LG A	557	2.55	1.26	49.42
B-LG B	554	2.44	1.69	69.31
Free Amino Acids, µg/mL				
Total FAA	715	64.12	22.41	34.95
Lys	686	4.52	4.26	94.35
Val	625	1.67	1.43	85.73
Glu	714	30.70	15.96	52.00
Gly	699	7.00	5.25	74.90
Asp	595	2.62	1.63	62.45
Arg	612	3.38	1.68	49.67
Ser	591	1.39	0.83	59.74

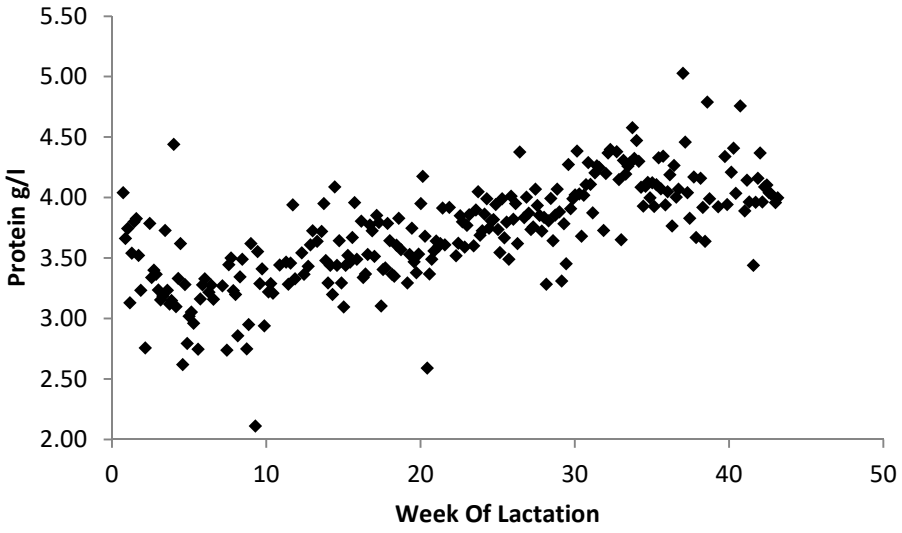


Figure 2.1 Trend in protein concentration across lactation.

2.4.1 Descriptive statistics

Mean gold standard values of all milk traits are summarised in **Table 2.1**. Mean values of α_{s1}- (13.92 g/L), α_{s2}- (3.62 g/L), β- (12.64 g/L), and κ-CN (5.92 g/L) in the present study were approximately in the ratio 3:1:3:1. Large differences were observed in the coefficient of variation (CV) across traits. The CV for protein fractions ranged from 20% (total-CN) to 69% (β-Lg-B).

The FAA present at the greatest concentration was Glu (mean = 30.70 μg/ml) but exhibited a large variability (standard deviation of 15.96 μg/ml), while Ser was present at the lowest concentration (mean = 1.39 μg/ml). The CV was generally large for all FAA, with a wide range from 35% (total FAA) to 94% (Lys).

Protein concentration decreased in early lactation and increased linearly across lactation thereafter (**Figure 2.1**). The lactation profile of (gold standard) total FAA (**Figure 2.2**) indicated that the concentration of total FAA were greatest in early and late lactation.

2.4.2 Protein prediction accuracy

Prediction accuracies achieved for cross validation and external validation are summarised in **Table 2.2.** The mean bias in prediction of protein fractions was not different from zero ($P > 0.05$). The number of factors included in the partial least squares prediction model varied from 4 (total-CN, α_{s1}-CN, β-CN, β-Lg-B) to 16 (β-Lg-A). The r_c between gold standard and MIRS-predicted protein fractions ranged from 0.43 (β-Lg-A) to 0.76 (Total Lg) and the greatest r_v values obtained for protein fractions were 0.67, 0.69 and 0.74 for β-CN, total β-Lg and total-CN, respectively. Total-CN also had the greatest RPD (1.49). The slope between the gold standard and MIRS-predicted values for protein fractions ranged from 0.76 (β-Lg-B) to 0.99 (κ-CN and β-CN). The average difference in r_v when undertaken across all breeds or within just the Holstein-Friesians (**Table 2.3**) varied from -0.08 (Arg) to 0.06 (α_{s1}-CN).

The r_v for the different proteins predicted from just protein content was on average 0.18 less than prediction of the same traits using MIRS. Expressing protein fractions as a percentage of total protein, accuracy of prediction was poorer than when proteins were expressed as grams per decilitre of milk (results not shown); the difference between r_v for traits when expressed as g/L milk compared to when expressed as a percentage of protein ranged from 0.01 (κ-CN) to 0.42 (α-Lac).

2.4.3 Prediction accuracy of free amino acids

Accuracy of the developed equations to predict FAA are summarised in **Table 2.3**. The number of factors included in the prediction model ranged from 9 (Ser and Arg) to 15 (Gly). Moderate prediction accuracy of FAA were achieved, particularly for Gly, Lys and Glu, with an r_c and r_v of 0.75 and 0.75 respectively, for Gly and an r_c of 0.68 and an r_v of 0.59, respectively for Glu; Gly also had the greatest RPD (1.38). Arg had the lowest r_v (0.26). The slope between the gold standard and MIRS-predicted values ranged from 0.67 (Ser) to 0.92 (Asp). FAA were on average over predicted ($P<0.05$).

2.4.4 Phenotypic correlations

Pearson correlations among the protein fractions are summarised in **Table 2.4.** Correlations among the gold standard proteins and among the MIRS-predicted protein fractions were all different ($P<0.05$) from zero. The correlations between gold standard total-CN and gold standard casein fractions ranged from 0.67 (α_{S2}-CN) to 0.92 (α_{S1}-CN and β-CN) and were similar to correlations between the MIRS-predicted total-CN and the

MIRS-predicted components of casein. However, the correlation between the gold standard values of α_{S1}-CN and α_{S2}-CN (0.53) was weaker than the respective correlation between the MIRS-predicted values (0.85).

The correlation between total-whey and total-Lg was 0.99 (gold standard values) and 0.94 (MIRS-predicted values). Similarly, the correlation between the gold standard α-La and total-Lg (0.47) was similar to the respective correlation between their MIRS-predicted values (0.48), whereas the correlations between the gold standard β-Lg-A and β-CN (0.36) and their corresponding MIRS-predicted values (0.79) differed.

Pearson correlations among gold standard FAA and among MIRS-predicted FAA are in **Table 2.5**. In general, the correlations among the gold standard FAA and the respective correlations among the MIRS-predicted FAA were in less agreement than the correlations among the gold standard or the MIRS-predicted protein fractions.

Pearson correlations among protein-related traits (i.e., MIRS-predicted protein, MIRS-predicted casein, gold standard protein fractions) and milk processing characteristics (i.e., RCT, k_{20}, a_{30} HCT and pH) in early (DIM<60) and late (DIM>180) lactation are in **Table 2.6**; all correlations were generally weak to moderate. Rennet coagulating time was positively associated with MIRS-predicted protein in early lactation (r=0.19), but was negatively correlated with MIRS-predicted protein in late lactation (r=-0.11), corresponding with the increase in protein concentration across lactation (**Figure 2.2**). In early lactation, RCT was negatively associated with MIRS-predicted casein (-0.21). Curd firming time was negatively correlated with the protein-related traits in both early and late lactation. The opposite was true for a_{30}, which was generally positively correlated with the protein-related traits in early and late lactation. Native pH was negatively correlated with gold standard protein fractions in early lactation, but was both negatively and positively correlated with gold standard protein fractions in late lactation. The correlations among HCT and β-LG in early and late lactation were -0.17 and 0.22, respectively. The Pearson correlations between HCT and κ-CN in both early and late lactation were not different from zero, (r = -0.05 and r = 0.08, respectively).

Table 2.2 Number of records (n), average number of factors (#L; rounded to the nearest whole number), root mean square error (RMSE), correlation coefficient between gold standard and predicted values in cross validation (r_c) and external validation (r_v), bias (SE in parentheses), slope (b:SE in parentheses) and ratio performance deviation (RPD) tested using the split sample cross validation and external validation.

	Cross Validation				External Validation				
Trait[1]	n	#L	RMSE	r_c	Bias(SE)	b(SE)	RMSE	r_v	RPD
Protein									
TotalCN	554	4	4.68	0.75	-0.0068(4.71)	0.98(0.04)	4.80	0.74	1.49
α_{S1}-CN	557	4	2.16	0.70	0.0057(2.23)	0.97(0.05)	1.26	0.66	1.35
α_{S2} CN	555	5	0.78	0.60	0.0072(0.80)	0.90(0.06)	1.99	0.66	1.22
β-CN	555	4	1.92	0.69	0.0008(1.99)	0.99(0.05)	2.37	0.67	1.33
κ-CN	556	6	1.25	0.67	-0.0037(1.26)	0.99(0.05)	0.81	0.56	1.33
Total Whey	549	6	1.17	0.76	0.0049(1.22)	0.87(0.04)	1.36	0.65	1.32
A-LA	551	8	0.26	0.58	0.0012(0.26)	0.88(0.06)	0.26	0.54	1.17
Total Lg	552	14	1.01	0.76	0.0015(1.06)	0.87(0.04)	1.20	0.69	1.38
B-LG A	557	16	1.14	0.43	0.0003(1.14)	0.94(0.09)	1.16	0.39	1.09
B-LG B	554	4	1.29	0.65	0.0016(1.50)	0.76(0.06)	1.39	0.44	1.15
Free Amino Acids									
Total FAA	715	12	16.29	0.69	-0.0487(17.87)	0.88(0.04)	17.79	0.61	1.26
Lys[1]	686	14	0.56	0.69	-0.6910(3.30)	0.89(0.04)	3.35	0.55	1.27
Val[1]	625	11	0.57	0.60	-0.3381(1.62)	0.76(0.04)	1.93	0.59	1.14
Glu[1]	714	13	0.41	0.68	-2.0689(13.22)	0.86(0.04)	0.46	0.59	1.20
Gly[1]	612	15	0.41	0.75	-0.4769(3.54)	0.91(0.04)	3.50	0.75	1.38
Asp[1]	595	10	0.55	0.58	-0.3744(1.67)	0.92(0.08)	1.66	0.44	1.15
Arg[1]	620	9	0.38	0.66	-0.2347(1.59)	0.91(0.05)	4.35	0.26	1.25
Ser[1]	591	9	0.48	0.51	-0.1460(0.79)	0.67(0.07)	1.22	0.42	1.07

[1]Traits were log transformed prior to analysis

Table 2.3 Number of records (n), root mean square error (RMSE), correlation coefficient between gold standard and predicted values in external validation (rv) for all breeds and a Holstein Friesians only, average difference in rv between all breeds and Holstein Friesians, tested using the split sample external validation.

	All Breeds			Holstein Friesian Only			
Trait[1]	n	RMSE	r_v	n	RMSE	r_v	Difference
Protein							
Total Casein	554	4.71	0.72	332	4.79	0.71	0.02
Alpha $_{S1}$ Casein	555	2.20	0.67	332	2.09	0.65	-0.02
Alpha $_{S2}$ Casein	555	0.81	0.53	332	0.81	0.46	0.06
Beta Casein	555	2.00	0.60	332	2.03	0.61	-0.01
Kappa Casein	556	1.27	0.63	333	1.24	0.63	0.00
Total Whey	547	1.38	0.66	326	1.41	0.66	-0.01
A Lactalbumin	551	0.27	0.51	329	0.28	0.49	0.01
Total Lg	553	1.24	0.68	330	1.21	0.70	-0.04
B Lg A	557	1.16	0.37	333	1.16	0.35	0.01
B Lg B	554	1.45	0.52	330	1.43	0.49	0.03
Free Amino Acids							
Total FAA	715	19.2	0.50	445	18.99	0.52	-0.03
Lys[2]	677	0.58	0.58	425	0.57	0.60	-0.02
Val[2]	641	1.63	0.56	393	1.55	0.55	0.01
Glu[2]	714	13.16	0.57	444	13.94	0.56	0.00
Gly[2]	698	3.57	0.73	435	3.59	0.74	-0.01
Asp[2]	603	0.63	0.47	386	0.62	0.49	-0.01
Arg[5]	632	4.33	0.27	387	3.27	0.39	-0.08
Ser[5]	591	0.77	0.39	360	0.73	0.43	-0.03

[1]α Lactalbumin = Alpha Lactalbumin, β Lg A = Beta Lactoglobulin A, β Lg B = Beta Lactoglobulin B, Total Lg = Total Lactoglobulin.

[2]Traits were log transformed prior to analysis

Correlations between gold standard FAA and milk processing characteristics in early (DIM<60) and late (DIM>180) lactation are in **Table 2.7**. Correlations were strongest among gold standard FAA and the milk processing characteristics in early lactation. Similar to the correlations with the gold standard FAA, RCT was positively associated with MIRS-predicted FAA late lactation (**Table 2.8**). However, a_{30} was positively associated with MIRS-predicted FAA in early lactation but was negatively associated with MIRS-predicted FAA in late lactation (**Table 2.8**). In early lactation, pH was negatively correlated with all gold standard FAA.

2.5 Discussion

The objective of the present study was to demonstrate the ability of MIRS to predict milk quality traits, including seven individual proteins and seven FAA. Predictions of these traits by MIRS could be of benefit to the dairy industry as MIRS is a low cost and efficient method for acquiring phenotypic information on milk quality using infrastructure and logistics for the acquisition of milk samples that already exists.

Limited studies exist evaluating the effectiveness of MIRS in predicting milk protein composition (Bonfatti et al., 2011a; Rutten et al., 2011; De Marchi et al., 2010) and no studies have evaluated the ability of the MIRS to predict FAA. Furthermore, comparison with other studies of MIRS-prediction accuracy for protein fractions is difficult due to differences in the dairy production system as well as methods of determining protein fractions used (i.e. different gold standard analyses, experimental design, different breeds, stages of lactations, parities, diets and milking times). For example, the level of crude protein in the diet affects the milk protein profile (Reid et al., 2015); animals in the present study were on a predominately grass based diet. To our knowledge, this is the first study to use data from a mainly grazed grass based production system to develop equations to predict protein composition and FAA from milk MIR.

Caseins constitute approximately 80% of milk protein and consist of α_{s1}-, α_{s2}, β-, and κ-CN fractions, typically in the ratio 3:1:3:1 (Farrell et al., 2004). Mean values of the respective caseins in the present study were consistent with this ratio. Mean values of 1.11 g/L and 4.97 g/L for both the gold standard and MIRS-predicted α-La and total-Lg were in the ratio 1:3, consistent with values documented by Farrell et al. (2004). Similarly, the ratio of total-CN to total whey was 6:1 irrespective of whether calculated using the gold standard or MIRS-predicted values. Multiple sampling dates, as well as variability attributed to the numerous research farms, breeds, parities and milking times

used to maximise the variation in the sample populations is a likely contributing factor to the greater coefficient of variation in protein fractions compared to other studies (De Marchi et al., 2010).

Glutamic acid was the FAA present in the greatest concentration in the milk. This conclusion was consistent with previous studies by Roucher et al. (2013), Lindmark-Mansson et al. (2003) and Sarwar et al. (1998), who also documented Glu to be one of the most abundant FAA in bovine milk. The FAA Asp, Arg and Ser were present in low concentrations in the present study.

The lactation profile of (gold standard) total FAA (**Figure 2.2**) indicated that the greatest concentration of FAA was during early and late lactation. This was similar to a finding by Ghadimi et al. (1963), who also documented variation in the concentration of FAA at different stages of lactation, with the greatest concentration of FAA present in the colostrum, and the least concentration in transitional milk.

2.5.1 Prediction of protein fractions

The r_c and r_v of total proteins (i.e., total-CN, total-whey and total β-lg) were predicted with greater accuracy than their components, which was probably attributed in part to their greater concentration in the milk. The ability to predict components in greater concentration in the milk corroborates the conclusion of Soyeurt et al. (2011), Soyeurt et al. (2006), and Rutten et al. (2009), who all attempted to predict milk fatty acid content using MIRS. Accuracy of prediction of protein fractions overall in the present study were consistent with those documented in other publications previously (Bonfatti et al. 2011a; Rutten et al. 2011; De Marchi et al. 2010). Differences among studies could be due to differences in the gold standard methods used. HPLC was used in this present study as well as in both the studies of Bonfatti et al. (2011a) and De Marchi et al. (2010), whereas Rutten et al. (2011) used capillary zone electrophoresis. The traits predicted with the poorest accuracy were β-lg A and β-lg B (r_v = 0.39 and r_v =0.44). This may be because the quantity of β-lg A and β-lg B are directly related to the milk protein variants of the cow; if cows are AA the content of β-lg B is 0 and if cows are BB the content of β-lg A is 0 (Ng-Kwai-Hang and Kim, 1996).

Table 2.4 Pearson correlations between gold standard (below diagonal) and mid-infrared spectroscopy predicted (above diagonal) protein composition traits.

	Prt %	Total-CN	α_{S1}-CN	α_{S2}-CN	β-CN	k-CN	Total-Whey	α-Lac	Total-Lg	β-Lg-A	β-Lg-B
Protein %	-	0.58	0.55	0.41	0.40	0.53	0.42	0.25	0.42	0.33	0.17
Total-CN	0.49	-	0.93	0.92	0.95	0.95	0.68	0.53	0.65	0.88	0.36
α_{S1}-CN	0.32	0.92	-	0.85	0.86	0.88	0.63	0.51	0.61	0.82	0.32
α_{S2}-CN	0.49	0.67	0.53	-	0.86	0.90	0.67	0.57	0.62	0.86	0.31
β-CN	0.45	0.92	0.79	0.52	-	0.87	0.63	0.58	0.60	0.79	0.33
k-CN	0.49	0.80	0.61	0.54	0.66	-	0.71	0.43	0.69	0.92	0.39
Total-Whey	-0.02	0.61	0.58	0.40	0.49	0.57	-	0.57	0.94	0.70	0.72
α-LA	-0.02	0.46	0.49	0.34	0.38	0.33	0.59	-	0.48	0.44	0.29
Total-Lg	0.02	0.58	0.54	0.38	0.47	0.57	0.99	0.47	-	0.67	0.77
β-Lg-A	0.48	0.43	0.43	0.25	0.36	0.40	0.39	0.34	0.37	-	0.33
β-Lg-B	-0.13	0.24	0.21	0.19	0.19	0.26	0.70	0.21	0.72	-0.38	-

[1] All correlations were different from zero ($P < 0.05$).

[2]Total Casein (Total CN), Alpha $_{S1}$ Casein (α-S1-CN), Alpha $_{S2}$ Casein (α-S1-CN), Beta Casein (β-Casein), Kappa Casein (k-CN), Total Whey, Beta Lactoglobulin A (β-Lg-A), Beta-Lactoglobulin-B (β-Lg-B) , Alpha Lactalbumin (α-LA) and Total Lactoglobulin (Total-lac).

Table 2.5 Pearson correlations between gold standard (below diagonal) and mid-infrared spectroscopy predicted (above diagonal) free amino acids.

Traits[1]	Lys	Val	Glu	Gly	Asp	Arg	Ser
Lys	-	0.69	-0.23	0.05	-0.18	0.69	0.31
Val	0.57	-	0.01	0.26	0.06	0.67	0.36
Glu	-0.10	0.2	-	0.38	0.70	-0.29	0.10
Gly	0.07	0.31	0.35	-	0.40	-0.1	0.30
Asp	-0.01	0.15	0.70	0.33	-	-0.4	-0.10
Arg	0.53	0.58	0.03	0.08	-0.06	-	0.39
Ser	0.19	0.35	0.34	0.43	0.14	0.43	-

[1]Traits were log transformed prior to analysis.
[2]Correlations $\leq \left| 0.07 \right|$ were not different from zero ($P > 0.05$).

A high RPD is advantageous; an RPD greater than two indicates the generated prediction could be used for analytical purposes (Williams et al., 2007). No RPD value greater than two was however, achieved in this present study. All protein fractions had an RPD between one and two in the present study and this finding is consistent with a previous study on milk protein fractions (Bonfatti et al., 2011a). According to Williams et al. (2007), a slope of the gold standard values on the MIRS-predicted values of a trait that deviates greatly from 1 (e.g., less than 0.85 and 1.15 or greater) will result in an unstable calibration, whereas a prediction equation with a slope between 0.95 and 1.05 will be more stable. The prediction models for four protein fractions (total-CN, α_{s1}-CN, β-CN and k-CN) had slopes between 0.95 and 1.00 in the present study. Protein fractions were however, on average under predicted. This could result in farmers being underpaid should a milk payment system on protein fractions be implemented.

A poorer accuracy of prediction was obtained when protein fractions were expressed as a percentage of MIRS-predicted protein or in milk; this was consistent with results from previous studies by Bonfatti et al. (2011a) on protein fractions and by Soyeurt et al. (2006) on fatty acid content. The poorer accuracy of prediction when protein fractions were expressed as a percentage could be explained by a variation in the protein fractions present in different milk samples. For example, two milk samples could have the same concentration of protein in the milk, but be made up of different protein fractions. Another possible explanation for poorer accuracy of prediction is that the protein content of milk was actually predicted and not the actual protein composition. Protein content and protein composition are highly correlated (expressed as g/L).

Table 2.6 Pearson correlations[1] between mid-infrared spectroscopy predicted protein, mid-infrared spectroscopy predicted casein, gold standard protein fractions and rennet coagulation time (RCT), curd-firming time (k_{20}), curd firmness (a_{30}), heat coagulation time (HCT) and pH across two stages of lactation.

	PRT	CN	Total-CN	α_{S1}-CN	α_{S2}-CN	β-CN	k-CN	Total-Whey	α-LA	Total-Lg	β-Lg-A	β-Lg-B
Early Lactation												
RCT	0.19	-0.21	0.20	0.14	0.23	0.20	0.17	0.09	0.06	0.08	0.24	-0.10
k_{20}	-0.09	-0.44	0.00	0.00	-0.05	0.05	-0.05	-0.12	-0.07	-0.11	0.03	-0.12
a_{30}	0.49	0.49	0.50	0.49	0.32	0.47	0.43	0.32	0.07	0.32	0.09	0.22
HCT	-0.2	-0.14	-0.10	-0.11	-0.14	-0.07	-0.05	-0.18	-0.10	-0.17	-0.02	-0.13
pH	-0.48	-0.45	-0.37	-0.47	-0.27	-0.25	-0.18	-0.26*	-0.25*	-0.22	-0.18	-0.06
Late Lactation												
RCT	-0.11	-0.17	-0.10	-0.16	0.05	-0.03	-0.13	-0.15	-0.08	-0.15	-0.06	-0.10
k_{20}	-0.27	-0.33	-0.31	-0.27	-0.16	-0.21	-0.35	-0.17	-0.05	-0.17	-0.11	-0.09
a_{30}	0.39	0.38	0.37	0.37	0.16	0.22	0.40	0.29	0.10	0.30	0.17	0.17
HCT	-0.24	-0.31	-0.08	-0.03	-0.16	-0.04	0.08	0.22	0.13	0.22	0.05	0.18
pH	-0.14	-0.12	0.01	-0.10	0.09	0.06	0.10	0.17*	0.22*	0.15	0.10	0.05*

*Correlations are significantly different to each other in early and late lactation ($P<0.01$)

[1]Protein (PRT), Casein (CN), Total Casein (Total-CN), Alpha $_{S1}$ Casein (α-S1-CN), Alpha $_{S2}$ Casein (α-S1-CN), Beta Casein (β-Casein), Kappa Casein (k-CN), Total Whey, Beta Lactoglobulin A (β-Lg-A), Beta Lactoglobulin B (β-Lg-B), Alpha Lactalbumin (α-LA) and Total Lactoglobulin (Total-LG).

[2]Correlations $\leq \left| 0.11 \right|$ were not different from zero ($P > 0.05$).

Table 2.7 Pearson correlations between gold standard free amino acids (FAA) and rennet coagulation time (RCT), curd-firming time (k_{20}), curd firmness (a_{30}), heat coagulation time (HCT) and pH; across two stages of lactation.

	Lys	Val	Glu	Gly	Asp	Arg	Ser	Total FAA
Early Lactation								
RCT	0.14	-0.02	0.19*	-0.05	0.04	0.16	-0.10*	0.05
k_{20}	-0.01	0.00	0.21	-0.11	0.17	0.07	-0.03	0.10
a_{30}	0.08	-0.02	0.26	0.32	0.19	0.00	0.24	0.25
HCT	-0.12	-0.01	0.06	-0.35	-0.06	0.11	-0.10	0.02
pH	-0.48*	-0.60*	-0.33*	-0.52*	-0.32*	-0.30*	-0.38*	-0.51*
Late Lactation								
RCT	0.06	0.20	0.40*	0.01	0.05	0.29	0.37*	0.28
k_{20}	-0.14	0.49	0.30	-0.01	0.08	0.10	0.23	0.18
a_{30}	0.06	-0.01	-0.34	-0.03	-0.11	-0.09	-0.26	-0.22
HCT	-0.23	0.03	0.08	-0.04	-0.13	0.02	0.08	0.01
pH	-0.09*	0.09*	0.09*	-0.06*	-0.01*	0.05*	0.27*	-0.02*

*Correlations are significantly different to each other in early and late lactation ($P<0.01$)

[1]Correlations $\leq |0.18|$ were not different from zero ($P > 0.05$)

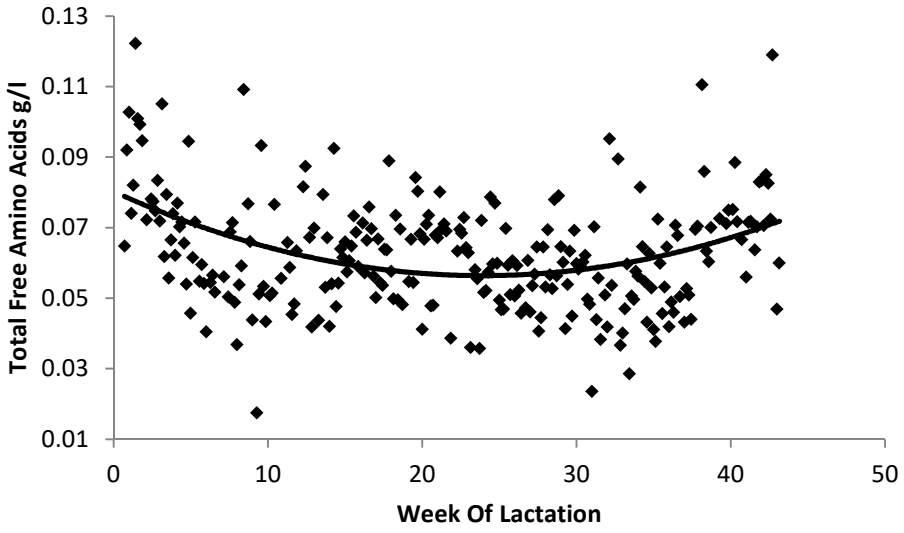


Figure 2.2 Trend in total free amino acids (i.e. the sum of gold standard Lys, Val, Glu, Gly, Asp, Arg, Ser) across lactation.

Nonetheless, exploiting the infrared spectrum in the prediction of milk protein composition generated superior prediction accuracy than when protein composition was predicted solely based on milk protein content suggest that the spectrum is in fact providing additional information in the prediction process.

2.5.2 Prediction of free amino acids

The present study is the first attempt to predict FAA in milk from MIRS. The moderate prediction accuracies for FAA achieved in the present study may be due to the low concentration of FAA present in the milk samples. The optimum number of factors included in the partial least squares prediction model for FAA was similar to previous studies (De Marchi et al., 2010; Soyeurt et al., 2011) for the prediction of casein fractions and fatty acid composition. Although FAA required a greater number of factors than for casein fraction predictions in this present study.

Glu which was predicted with moderate accuracy from the MIR may be important for infant formula production as the sum of Glu and Gln represents 50% of the total FAA in human milk (Agostoni et al., 2000). The prediction ability for all FAA was too poor for industrial use. The prediction models for all FAA had slopes of the gold standard values on the MIRS-predicted values between 0.85 and 1, with the exception of Val (0.76) and Ser (0.67) which had slopes of the gold standard values on the MIRS-predicted values of <0.85 and therefore they may have unstable calibrations (Williams et al., 2007).

Since protein and FAA are correlated; the MIRS could be indirectly predicting the FAA, by predicting the protein content of the milk.

2.5.3. Phenotypic correlations

The correlations among the gold standard traits were comparable to the correlations among the corresponding MIRS-predicted traits for the majority of protein fraction traits. This could be due to the moderately accurate predictions, which yielded MIRS-predicted protein fraction values similar to gold standard protein fraction values. The correlation between gold standard β-Lg-A and gold standard β-Lg-B (-0.38) was not in agreement with the correlation between the MIRS-predicted β-Lg-A and the MIRS-predicted β-Lg-B (0.33), as these protein

fractions β-Lg-A (r_v =0.39) and β-Lg-B (r_v =0.44) were poorly predicted from the MIR. The poor prediction of β-Lg-A and β-Lg-B may also contribute to the large difference in the correlations between the gold standard and MIRS-predicted values of these two traits with total-CN.

Overall, the correlations among the gold standard FAA and the respective correlations among the MIRS-predicted FAA were in less agreement than the correlations among the gold standard and the MIRS-predicted protein fractions this may be due to poorer accuracy of prediction for FAA.

Previous studies have shown that RCT, k_{20} and pH are all positively correlated and these traits are all negatively correlated with both a_{30} and HCT (Ikonen et al., 2004; Cassandro et al., 2008; Visentin et al., 2015). Therefore, if RCT has a positive correlation with protein; k_{20} and pH should also have positive correlations with protein and both a_{30} and HCT should be negatively correlated with protein. This study estimated RCT was positively correlated with MIRS-predicted protein and the majority of gold standard protein fractions in early lactation, but in late lactation RCT was negatively correlated with MIRS-predicted protein and gold standard protein fractions. The correlations between RCT, MIRS-predicted protein and gold standard protein fractions could be explained by the increase in protein content throughout lactation (**Figure 2.2**), because as protein content increases, RCT decreases (Visentin et al., 2015) and also the FAA concentration decreases. Throughout lactation, k_{20} had negative associations with protein-related traits but had positive associations with the majority of gold standard FAA; this could be due to protein hydrolysis, which releases FAA into the milk. The negative associations of k_{20} with protein-related traits and the positive associations of k_{20} with the majority of gold standard FAA were also in accordance with the positive correlation among RCT and k_{20}. RCT was negatively correlated with a_{30}; therefore the positive correlations between a_{30} and protein-related traits in both early and late lactations and the negative associations with gold standard FAA in mid lactation were expected. Negative correlations were demonstrated between pH and both MIRS-predicted casein and gold standard protein fractions in early lactation, but both negative and positive correlations were demonstrated between pH and these traits in late lactation. These correlations may be explained by Vasbinder et al. (2003), who showed that a small

change in pH had a large effect on whey protein denaturation and gelation properties of milk. Milk proteins, in particular β-LG and κ-CN are known to have an effect on milk processing characteristics, such as HCT (Singh, 2004). However no strong correlations were demonstrated between these proteins and milk processing characteristics in this present study.

Table 2.8 Pearson correlations between mid-infrared predicted free amino acids (FAA) and rennet coagulation time (RCT), curd-firming time (k_{20}), curd firmness (a_{30}), heat coagulation time (HCT) and pH; across two stages of lactation.

	Lys	Val	Glu	Gly	Asp	Arg	Ser	Total FAA
Early Lactation								
RCT	-0.10*	-0.19*	0.24	-0.04	-0.14	-0.01	-0.16	-0.10
k_{20}	0.01	-0.04	0.13	-0.06	0.00	-0.08	-0.20	-0.03
a_{30}	0.20*	0.23*	-0.21	0.10	0.00	-0.11	-0.04	0.13
HCT	0.13	-0.02	0.15	-0.04	0.29	-0.10	0.17	0.13
pH	-0.15*	-0.18	-0.01	-0.13	-0.12	-0.22	-0.13	-0.24*
Late Lactation								
RCT	0.34*	0.21*	0.08	0.23	0.06	0.02	0.02	0.37
k_{20}	0.17	0.10	0.17	0.27	0.12	0.01	0.03	0.27
a_{30}	-0.18*	-0.09*	-0.17	-0.24	-0.11	0.01	-0.08	-0.26
HCT	-0.14	-0.08	0.05	-0.01	-0.10*	0.24*	0.06	-0.07
pH	0.23*	-0.08	-0.06	0.07	0.03	-0.02	0.00	0.13*

* Correlations are significantly different to each other in early and late lactation ($P<0.01$).

2.6 Conclusions

Findings from this study indicate that MIRS is useful to routinely and efficiently measure milk quality traits such as protein fractions and some FAA at a population level. Prediction of these traits by MIRS could play an important role in selective breeding and therefore be of benefit to the dairy and breeding industry worldwide, allowing for the more accurate selection of milk for human consumption, infant milk formula and cheese production. Further research is required to quantify genetic correlations between protein fractions and FAA and to estimate the genetic variance of these traits which will indicate the usefulness of the developed MIRS models for practical animal breeding purposes.

2.7 Acknowledgements

Funding for this work was received from the Irish Department of Agriculture, Food and the Marine, Research Stimulus Fund project 11/SF/311, Breed Quality. The authors gratefully acknowledge the laboratory assistance of Bernard Corrigan (Teagasc, Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland).

CHAPTER 3

Effectiveness of mid-infrared spectroscopy to predict the colour of bovine milk and the relationship between milk colour and traditional milk quality traits

A. McDermott*†, G. Visentin *† , S. McParland*, D.P. Berry*, M.A. Fenelon‡, and M. De Marchi†[1]

*Teagasc Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

†Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Italy

‡Teagasc, Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland.

J. Dairy Sci. 99:1–7

3.1 Abstract

The colour of milk impacts the subsequent colour features of the resulting dairy products; milk colour is also related to milk fat concentration. The objective of the present study was to quantify the ability of mid-infrared spectroscopy (MIRS) to predict colour-related traits in milk samples and to estimate the correlations between these colour-related characteristics and traditional milk quality traits. Mid-infrared spectral data were available on 601 milk samples from 529 cows, all of which had corresponding gold standard milk colour measures determined using a Chroma Meter; milk colour was expressed using the CIELAB uniform colour space. Separate prediction equations were developed for each of the three colour parameters (L* = lightness, a* = greenness b*=yellowness) using partial least squares regression. Accuracy of prediction was determined using both cross validation on a calibration data set (n=422 to 457 samples) and external validation on a data set of 144 to 152 samples. Moderate accuracy of prediction was achieved for the b* index (coefficient of correlation for external validation = 0.72), although poor predictive ability was obtained for both a* and L* indices (coefficient of correlation for external validation of 0.30 and 0.55, respectively). The linear regression coefficient of the gold standard values on the respective MIRS-predicted values of a*, L*, and b* was 0.81, 0.88, and 0.96, respectively; only the regression coefficient on L* was different ($P<0.05$) from one. The mean bias of prediction (i.e., the average difference between the MIRS-predicted values and gold standard values in external validation) was not different from zero ($P > 0.05$) for any of three parameters evaluated. A moderate correlation (0.56) existed between the MIRS-predicted L* and b* indices, both of which were weakly correlated with the a* index. Milk fat, protein and casein were moderately correlated with both the gold standard and MIRS-predicted values for b*. Results from the present study indicate that MIRS data provides an efficient, low-cost, screening method to determine the b* colour of milk at a population level.

3.2 Introduction

Product colour is one of the primary factors considered by consumers when making purchasing decisions, as it is often an indicator of ripeness, freshness, food safety and attractiveness in the food industry (Hutchings 1994). It is well known that milk colour influences the colour features of the subsequent dairy products, while also being related to the fat content of the milk (Winkelman et al., 1999). Differences in milk colour can also be related to the presence of abnormalities in milk; for example, mastitis attributable to *Streptococcus esculine* infection causes milk to have a more reddish/yellowish colour while mastitis due to *Streptococcus dysgalactiae* also leads to a change in milk colour (Espada et al., 2002 and Vijverberg, 2002).

The white colour of milk is a function of the milk's physical structure; the dispersion of both casein micelles and fat globules in the milk is responsible for the diffusion of incident light and is related to lightness (L*) (Raty et al., 1999). The natural pigmentations from carotenoids, protein and riboflavin are also associated with the white colour of milk. Milk with a low carotenoid content, high protein and high riboflavin tends to be whiter (Solah et al., 2007), or in other words have a greater L* index value.

The yellow colour (yellowness index; b*) of bovine milk is closely related to the level of β-carotene and fat content; a greater milk fat and β-carotene content results in an incremental increase to the b* index of milk, hence the milk will have a more yellow colour. Feeding and selective breeding of cows may be used to alter the carotenoid level and thus colour of dairy products (Norieze et al., 2006b). Cows fed grass silage tend to produce milk with yellower fat and greater β-carotene content, than milk produced by cows on a hay diet (Noziere et al., 2006a; Calderon et al., 2007). Breeds of cows, such as Jerseys, that produce milk with a greater carotenoid and fat content, produce more yellow colour milk than breeds such as Holstein-Friesians (Winkelman et al., 1999). There is a minimal loss of carotenoids from milk when transferred into butter and cheese therefore also contributing to the yellow colouration of these dairy products.

Yellower dairy products may be considered favourable or unfavourable depending on the target market. For example, yellower products are considered an unfavourable attribute in Middle Eastern dairy markets (Keen and Wilson, 1992).

However in Europe, a yellower colour is favourable in high fat dairy products such as butter and full fat cheeses (Hutchings 1994, Casalis et al., 1972).

As the gold standard methods for the determination of milk colour (i.e., Chroma Meter (Minolta, Osaka, Japan) or a NH310 Colour Meter Milk (Shenzhen 3NH Technology Co. Ltd, China) or for the determination of milk carotenoid content, can be relatively costly and also requires sub-sampling of milk for analysis, the use of an analytical system already in place (e.g. mid-infrared spectroscopy) to determine milk colour may be more logical. Mid–infrared spectroscopy (MIRS) is currently used by milk recording organizations worldwide to predict milk fat, protein, casein and lactose concentration and has recently been used to predict more detailed milk composition traits (De Marchi et al., 2014) or animal traits (McParland et al., 2014). The use of MIRS to predict novel milk quality traits is therefore appealing since the MIR spectrum is available at a negligible additional cost and may be undertaken as part of the routine quantification of other components in milk. Nevertheless, to our knowledge, no study has attempted to evaluate the potential of MIRS to predict milk colour traits.

The aim of the present study was to evaluate the ability of MIRS to predict milk colour-related traits and to estimate the correlations between these milk colour traits and a selection of traditional milk quality traits.

3.3 Materials and methods

3.3.1 Milk sample collection

Between August 2013 and August 2014, inclusive, 730 milk samples from 621 cows were obtained from seven research farms operated by the Teagasc Animal and Grassland Research and Innovation Centre (Moorepark, Fermoy, Co.Cork, Ireland). Milk composition was recorded weekly using a MilkoScan FT6000 (Foss Electronic A/S, Hillerød, Denmark) and the resulting spectrum, containing 1,060 transmittance data in the mid-infrared region between 900 and 5,000 cm^{-1}, was stored. Following MIRS analysis, the milk samples were stored at 4 °C for further analysis. Samples were selected to maximize diversity of breed [Holstein-Friesian (n=454), Jersey (n=117), Norwegian Red (n=15), and crossbreds (n=144)], stage of lactation (5 to 375 days in milk), milking time (i.e. AM or PM milking), and parity (1 to 11). Samples with preservative added

(n=129) were not considered in the present study for the determination of milk colour. The final data set used in the present study comprised of 601 milk samples, 461 of which were from spring-calving cows fed a predominantly grazed grass based diet and the remaining 140 samples were from autumn-calving cows fed a total mixed ration diet in the early stages of lactation.

3.3.2 Gold standard milk colour determination

Milk colour was measured using a Chroma Meter CR400 (Konica Minolta Sensing Europe, Edisonbaan 14-F, NE) with a closed cone, set on the L* a* b* system and the Chroma meter was calibrated on a white tile. A 10 ml sub-sample of each milk sample was measured in a cuvette and expressed using the CIE-L* a* b* uniform colour space (CIE-Lab 1976). The CIE-L* a* b* plots the colour co-ordinates in a uniform colour space, which has an L* a* and b* axis, where L* = lightness [on a scale from 0 to 100 where 0 = black and 100 = white], a* [where - a* has a green colour and + a* has a red colour] and b* [where – b∗ has a blue colour and + b∗ has a yellow colour]. The more different from zero or the greater the absolute value is, the stronger the colour (i.e. a sample with an absolute value close to zero has a lighter colour than a sample with an absolute value close to one hundred).

3.3.3 Data analysis

Outlier samples were considered to be samples with a gold standard value >3 from the mean. No L* or b* indices outliers were removed, but 28 outliers were removed based on the a* index. All three milk colour traits were normally distributed. Descriptive statistics were calculated within the Holstein-Friesian and Jersey breeds separately, as well as across all breeds combined and within season. The differences between the means of Holstein Friesian and Jersey cows and between the means of autumn and spring cows were derived using ANOVA in Microsoft Excel. Spectral data were transformed from transmittance to linear absorbance using a logarithmic transformation of the reciprocal of the wavelength values (Soyeurt et al., 2011). Prediction models were developed using untreated spectra. Mid-infrared spectroscopy models were developed to predict each colour trait separately using partial least squares regression (Proc PLS; SAS Institute Inc., Cary, NC). Spectral regions considered were from 926-1,580 cm^{-1}, 1,717-2,986

cm^{-1} and 3,696-3,808 cm^{-1}, determined based on the observed loadings for each wavelength from preliminary analyses.

Accuracy of the prediction equations was determined using external validation, whereby 25% of data was excluded from equation calibration; these data were used in the independent validation data set. This procedure was repeated 3 times, using a different 25% of the data in the validation data set each time. Samples were selected for the validation data set to represent similar variation to that present in the gold standard data in the calibration data set used to develop the prediction equations. For each prediction model, the data set was sorted by the trait of interest. The first sample and every fourth sample thereafter were included in the validation data set for the first iteration; for the second iteration the second sample and every fourth sample thereafter was chosen for the validation data set, with a similar procedure used for the third and fourth iteration. No cow was represented in both the calibration data set and validation data set in a given iteration.

The criteria used to determine the accuracy of the MIRS predictive models were the coefficient of correlation of cross validation (r_c) and external validation (r_v), the root mean square error of cross validation (RMSEc) and external validation (RMSEv), the linear regression coefficient (b) of the MIRS-predicted values on the gold standard values of each trait, and the mean bias of prediction; the mean bias of prediction was calculated as the mean difference between the MIRS-predicted values and gold standard values in external validation The ratio performance deviation (RPD), which is the ratio of the standard deviation of each trait to standard error of prediction, was also used as a measure of model predictive ability. Four validation datasets were created and then appended onto each other and the r_c, r_v, RMSE, b (SE), and bias (SE) was calculated based on all four iterations combined. The average number of factors (#L) used to build the prediction equations, was the average number from all four iterations rounded to the nearest whole number. Validation was performed within the Holstein-Friesian and Jersey breeds separately, as well as across all breeds combined. The Fischer's r to z transformation was performed to determine if the accuracy of prediction differed between breed populations.

Furthermore, the L*, a* ad b* indices were also predicted in external validation using the regression coefficients on milk fat, protein and casein content estimated from the calibration dataset.

Pearson correlations among the gold standard milk colour traits, MIRS-predicted milk colour traits and MIRS-predicted fat, protein and casein predicted using the FOSS equations (MilkoScan™ FT+) were also estimated.

3.4 Results

3.4.1 Descriptive statistics

Descriptive statistics of the milk colour traits are in **Table 3.1**; average values of a*, b* and L* colour indices were -3.88, 8.09, and 81.57, respectively. Jersey cows had a greater (P<0.01) mean value for the yellow colour of milk (b* = 10.03) than the Holstein-Friesian cows (b* =7.48) and their milk also had a greater fat content. Spring calving cows had a greater (P<0.01) b* index than autumn calving cows (results not shown). The coefficient of variation was 2.24%, 13.65%, and 36.34% for L*, a* and b*, respectively. **(Table 3.1)**

Table 3.1 Mean, standard deviation (SD), coefficient of variation (CV), minimum (Min) and maximum (Max) for the gold standard colour indices (L* = lightness; a* = greenness; b* = yellowness) in the entire dataset.

		Cross validation		External Validation				
Trait	# L	RMSE	r_c	Bias(SE)	b(SE)	RMSE	r_v	RPD
L*	15	1.46	0.63	0.02(1.57)	0.88(0.05)	1.57	0.55	1.20
a*	9	0.51	0.37	-0.002(0.52)	0.81(0.11)	0.52	0.30	1.05
b*	19	1.97	0.74	-0.005(2.03)	0.96(0.04)	2.03	0.72	1.45

[1]#L = average number of factors rounded to the nearest whole number RMSE = root mean square error; r_c = correlation between true and predicted values in cross validation; r_v = correlation between true and predicted values in external validation; b = linear regression coefficient of predicted values on the gold standard values of each trait SE = standard error; RPD = ratio performance deviation.

[1]The greater the absolute value is, the stronger the colour.

Table 3.2 Fit statistics[1] for the cross- and external- validation of prediction equations for colour indices (L* = lightness; a* = greenness; b* = yellowness).

	Cross validation			External validation				
Trait	# L	RMSE	r_c	Bias(SE)	b(SE)	RMSE	r_v	RPD
L*	15	1.46	0.63	0.02(1.57)	0.88(0.05)	1.57	0.55	1.20
a*	9	0.51	0.37	-0.002(0.52)	0.81(0.11)	0.52	0.30	1.05
b*	19	1.97	0.74	-0.005(2.03)	0.96(0.04)	2.03	0.72	1.45

[1]#L = average number of factors rounded to the nearest whole number RMSE = root mean square error; r_c = correlation between true and predicted values in cross validation; r_v = correlation between true and predicted values in external validation; b = linear regression coefficient of predicted values on the gold standard values of each trait SE = standard error; RPD = ratio performance deviation.

3.4.2 Colour prediction accuracy

Moderate accuracy of prediction was obtained for the b* index (r_v = 0.72; **Table 3.2**) while poor prediction accuracy of prediction was obtained for both the a* and L* indices (r_v = 0.30 and r_v = 0.55, respectively). The accuracy of predicting the a* index was greater (P<0.05) for the Jersey population (r_v = 0.59) compared to external validation in just the Holstein-Friesian population (r_v = 0.09) (**Table 3.3).** The accuracy of predicting the L* index was greater (P<0.05) for the Holstein-Friesian population (r_v = 0.60) and the Jersey population only (r_v = 0.73) compared to when all the breeds combined (r_v = 0.55) were included in external validation. The accuracy of predicting the b* index was greater (P<0.05) when all breeds combined were used in external validation (r_v = 0.72) compared to when just the Holstein-Friesian (r_v = 0.64) or Jersey (r_v =0.66) was used. RPD values for the three milk colour traits were less than 2. The linear regression coefficients of the MIRS-predicted values on the gold standard values for a*, L* and b* prediction models were 0.81, 0.88, 0.96, respectively; only the linear regression coefficient of L* was different to one (P<0.05). The bias of the prediction models, which is the average difference between MIRS-predicted values and gold standard values in external validation, was not different from zero (P > 0.05) for any of the three milk colour parameters.

Table 3.3 Number of records (n), root mean square error (RMSE), correlation coefficient between gold standard and MIRS-predicted values in external validation (r_v) for all breeds, Holstein-Friesians only and Jerseys only, tested using the split sample external validation.

Trait	n	Mean	SD	RMSE	r_v
All Breeds					
L*	120	81.57	1.83	1.57	0.55[a]
a*	143	-3.88	0.53	0.52	0.30[a]
b*	120	8.09	2.94	2.03	0.72[a]
Holstein-Friesians Only					
L*	98	81.57	1.91	1.55	0.60[b]
a*	98	-3.79	0.61	0.45	0.09[b]
b*	96	7.48	2.78	2.21	0.64[a]
Jerseys Only					
L*	94	81.40	1.74	1.20	0.73[ab]
a*	74	-4.25	1.34	0.48	0.59[a]
b*	94	10.03	3.04	2.30	0.66[a]

[ab] Coefficient of correlations within trait with different superscripts differ ($P < 0.05$) from each other.

When milk fat, protein and casein concentration combined were used as a predictor of the a*, L* and b* indices, r_v values (0.18, 0.48 and 0.60 for a*, L* and b*, respectively) were numerically lower ($P > 0.05$) than when predictions were based on the MIRS (0.30, 0.55 and 0.72 for a*, L* and b*, respectively).

3.4.3 Phenotypic correlations

A moderate correlation of 0.56 existed between the MIRS-predicted colour traits L* and b*, whereas the colour parameter a* was only weakly correlated with the other two colour traits (**Table 3.4**). A moderate correlation (0.65) existed between the gold standard colour traits L* and b*, whereas both the gold standard and MIRS-predicted colour parameter a* were only weakly correlated with the other two colour traits (**Table 3.4**). A negative correlation existed between the gold standard b* and a* indexes (-0.17), as well as between the MIRS-predicted b* and a * indexes (-0.04). The colour traits b* and L* were moderately correlated to MIRS-predicted milk fat, protein and casein content (**Table 3.4**). The correlation between the gold standard b* and MIRS-predicted fat (0.65) and between the MIRS-predicted b* and MIRS-predicted fat (0.59) were similar. No strong correlations existed between the gold standard and MIRS

predicted colour traits L* a* and b* with MIRS-predicted lactose. MIRS-predicted protein and CN were strongly correlated with CN constituting 77% of the variability in protein content.

Table 3.4 Pearson correlations[1] among the gold standard (below diagonal) and MIRS-predicted (above diagonal) colour indices (L* = lightness; a* = greenness; b* = yellowness) and MIRS-predicted traditional milk quality traits.

	L*	a*	b*	Fat, %	PRT, %	CN, %	Lactose, %
L*	-	0.35	0.56	0.38	0.36	0.39	0.01
a*	0.01	-	-0.04	-0.03	-0.17	-0.21	-0.07
b*	0.65	-0.17	-	0.59	0.48	0.45	-0.21
Fat %	0.49	-0.11	0.65	-	0.41	0.42	-0.05
PRT, %	0.42	-0.33	0.50	0.41	-	0.88	-0.06
CN, %	0.47	-0.32	0.47	0.42	0.88	-	0.13
Lactose, %	-0.01	-0.07	-0.29	-0.05	-0.06	0.13	-

[1]Correlations <|0.07| were not different from zero.

[2] PRT=protein CN=casein

3.5 Discussion

Feeding management could be an effective short term method of increasing the yellow colour in milk (Noziere et al., 2006a; Calderon et al., 2007), but selective breeding could be used as a long term strategy. Moreover, management strategies to alter milk colour may not always be feasible (e.g., feeding high producing cows forage diets). Genetic selection, however, is cumulative and permanent but more importantly is amenable for implementation globally. Successful breeding programs are nonetheless predicated on access to large quantities of individual animal level information from which to generate estimated breeding values; this information should ideally be available at low cost. This was the motivation of the present study to evaluate the potential of the routinely used MIRS to predict milk colour.

The mean absolute colour values in the present study were greater than a study by Ordolff (2006) who compared milk colour across a range of different somatic cell counts from 15 dairy cows in Germany. The b* values documented by Ordolff (2006) had negative values indicating the milk had a blue colour; this was expected as the milk had a very low fat content compared to the milk samples

analysed in the present study. The greater fat content of milk used in the present study may be because the majority of the milk samples were from cows fed a grass based diet. The yellow colour of milk is positively associated with greater Lucerne (Larsen et al., 2013). Cows fed grass silage tend to produce milk with a greater fat and β-carotene content and therefore a more yellow colour, than milk produced by cows on a hay diet (Noziere et al., 2006a; Calderon et al., 2007). A study by Phillips et al. (1995) that characterised low fat milk, calculated L- a- and b-values based on Illuminate A from a Macbeth Colour-Eye Spectrophotometer (Kollmorgen Instruments Corp., Newburgh, NY). The mean values for a, b and L observed by Phillips et al. (1995), at 2% fat content were -3.74, 2.99 and 81.11, respectively. The b value documented by Phillips et al. (1995) was also lower than the b value in the present study, where milk had a greater fat content.

Jersey cows had a greater ($P<0.01$) mean value for the yellow colour of milk ($b^* = 10.03$) than the Holstein-Friesian cows ($b^* =7.48$) and their milk also had a greater fat content.

Spring calving cows had a greater ($P<0.01$) b^* index than autumn calving cows. This is expected as spring-calving cows were fed a grass based diet, and therefore would have greater carotenoid and fat level in their milk, in comparison to autumn calving cows that were kept indoors and fed hay or silage (Noziere et al., 2006a; Calderon et al., 2007). Although bovine milk colour traits are related to the carotenoid level in milk and the transfer of these carotenoids from the blood to the milk (Nozière et al., 2006b; Gross et al., 2014), no information was available in the present study on the carotenoid content of the milk.

3.5.1 Milk mid-infrared spectroscopy prediction of Colour

Because the ratio performance deviation values for the three milk colour traits were less than 2 in the present study, the MIRS models should not be used for analytical purposes. According to Williams et al. (2007), the prediction model for a^* is unstable as the slope deviates greatly from 1 (e.g., less than 0.85 or greater than 1.15), whereas the prediction equation for b^* is expected to be stable since the linear regression coefficient of the true on predicted b^* was between 0.95 and 1.05. The prediction of milk colour indices by MIRS is not a consequence of direct prediction because the visible region of the electromagnetic spectrum (i.e., 350-800 nm) is not part of the MIRS spectrum (De Marchi et al.,

2014). As in the case of several innovative milk traits (e.g. milk coagulation traits; residual feed intake; methane emission), the prediction of colour traits in the present study is most likely from indirect prediction of other components, for example fat content; the correlation between the gold standard b* and fat content was 0.65 (**Table 3.4**). Mid-infrared spectroscopy is known to be able to accurately predict milk fat, protein, casein and free fatty acids and is currently used by milk recording organizations worldwide. The greater accuracy of MIRS prediction model for b* was also confirmed by the fewer number of model factors (9, 15 and 19 for a*, L* and b*, respectively). With partial least square regression, the pattern among wavelengths is reduced into fewer variables called loadings; each loading explains a part of the total variance and therefore the fewer the loading factors used the more robust the prediction equation is likely to be (De Marchi et al., 2013). The loadings also depict the molecular basis of the MIRS prediction, as peaks close to certain wavelengths are related to certain chemical bonds. Several spectral regions contributed to the prediction of the three colour traits in the present study, and these were primarily the regions associated with lipids (2,935, 2,839, 1,763 cm^{-1}), followed by the peaks at 968, 1,146, 1,180, 1,331 and 1,466 cm^{-1} which attributed to C-O, C-C, O-C-H, C-C-H and C-O-H bending (De Marchi et al., 2013). This helps verify that the ability of MIRS to predict the b* colour of milk may be (in part) an artefact of the MIRS detecting the milk fat concentration in the milk. Nonetheless, when the milk fat concentration alone was used as a predictor of the b* index, an r_v of only 0.33 was obtained, demonstrating that fat concentration alone is not predicting b* but other components represented in the MIRS are also contributing to the prediction of b*. Nonetheless, further studies should consider measuring the yellow milk colour independent of the fat content. This could be achieved by gravity separation of a portion of the milk samples that have a high and low yellow colour. The gravity cream could be added back in different combinations with gravity skim with low or high yellow colour and gravity cream with low or high yellow colour. This would enable the statistical model to better focus on information in the spectra that is related to colour and reduce the confounding effect of fat concentration.

Although no study has evaluated the potential of MIRS in predicting milk colour traits, several studies have investigated the use of near infrared spectroscopy (NIRS) to predict colour in other dairy products, especially cheese.

Lucas et al. (2008) reported robust NIRS prediction models for the colour of fresh cheeses (n = 445), with r_v values as high 0.96 for the colour traits a* and b*. Other studies measured colour traits in foods such as meat (De Marchi et al., 2011) and wine (Urbano-Cuadrado et al., 2004). De Marchi et al. (2011) obtained r_v values of 0.60, 0.82 and 0.88 for L* a* and b*, respectively in meat. Moreover, infrared spectroscopy has been proposed by Ordolff (2006) for the detection of abnormal milk colour in the blue region; for example, milk with reddish colour indicates contamination with blood that might be related to udder infection or teat injuries (Hettinga et al., 2008).

3.5.2 Phenotypic Correlations

Protein and casein were strongly correlated with each other (0.88), and therefore similar correlations between the colour traits (L* a* b*) with both protein and casein were expected. Correlations between milk fat and b* and between milk protein and a* in the present study were in agreement with those reported by Solah et al. (2007), who also reported significant correlations between these traits. Moreover, the correlation between MIRS-predicted lactose content and MIRS-predicted b* reported in the present study (-0.21) was weaker than that reported by Gross et al. (2014) in colostrum (-0.44).

3.6 Conclusions

This study demonstrates that MIRS data could be used as a screening tool to efficiently determine the b* colour of milk at a population level, providing a useful tool for the dairy industry and aiding in feeding management and selective breeding. One potential use of the equations developed in the present study is in milk cooperatives selecting specific farms that produce milk with certain milk colour characteristics which could demand a premium milk price. The MIRS prediction equations could be used to routinely monitor herd average milk colour for recruiting or expelling producers from this scheme. Further investigation is however, required to estimate the genetic variance of milk colour, which will indicate the usefulness of the developed MIRS models for practical animal breeding purposes.

3.7 Acknowledgements

Funding for this work was received from the Irish Department of Agriculture, Food and the Marine, Research Stimulus Fund project 11/SF/311, Breed Quality. The authors gratefully acknowledge the laboratory assistance of Owen Kenny (Teagasc, Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland).

CHAPTER 4

Cow and environmental factors associated with proteins fractions and free amino acids predicted using mid-infrared spectroscopy in bovine milk

A. McDermott*†, M. De Marchi†, D.P. Berry*, G. Visentin*†, M.A. Fenelon‡, N. Lopez-Villalobos§, and S. McParland*[1]

*Teagasc Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

†Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Italy

‡Teagasc, Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland.

§Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand.

Accepted for publication in the Journal of Dairy Science

4.1 Abstract

The objective of the present study was to identify the factors associated with both the protein composition and free amino acid (FAA) composition of bovine milk predicted using mid-infrared spectroscopy (MIRS). Milk samples were available from seven research herds and 69 commercial herds. The spectral data from the research herds comprised of 94,286 separate mornings and evening milk samples; the spectral data from the commercial herds comprised of 40,260 milk samples representing a composite sample of both the morning and evening milking. Mid-infrared spectroscopy prediction models developed in a previous study were applied to all spectra. Factors associated with the MIR-predicted protein and FAA composition were quantified using linear mixed models. Factors considered in the model included the fixed effects of calendar month of the test, milking time (i.e., AM, PM or both combined), parity (1, 2, 3, 4, 5 and ≥6), stage of lactation, the interaction between parity and stage of lactation, breed proportion of the cow (Friesian, Jersey, Norwegian Red, Montbelliarde, and other) and both the general heterosis and recombination coefficients of the cow. Contemporary group as well as both a within and across lactation permanent environmental effect were included in all models as random effects. Total proteins (i.e., total CN, total whey, and total β-LG) and protein fractions (with the exception of α-LA) decreased post-calving until 36-65 DIM and increased thereafter. After adjusting the statistical model for differences in crude protein content and milk yield separately, irrespective of stage of lactation, younger animals produced more total proteins (i.e., total CN, total whey, and total β-LG) as well as more total FAA, Glu, and Asp than their older contemporaries. The concentration of all protein fractions (except β-CN) in milk was greatest in the evening milk, even after adjusting for differences in the crude protein content of the milk. Relative to a purebred Holstein cow, Jersey cows, on average, produced a greater concentration of all CN fractions but less total FAA, Glu, Gly, Asp, and Val in milk. Relative to their respective purebred parental average, first-cross cows produced more total CN and more β-CN. Results from the present study indicate that many cow-level factors, as well as other factors, are associated with protein composition and FAA composition of bovine milk.

4.2 Introduction

Both total protein content, and its composition, in bovine milk are among the most important milk characteristics for the dairy industry. In Europe, milk processors pay a greater premium for milk exceeding a threshold protein content than for milk exceeding a threshold fat content (Shalloo and Geary, 2011). Bovine milk generally consists of about 3.3% protein, of which 78% is comprised of casein (CN), 17-18% is comprised of whey protein and the remaining 4-5% of non-protein nitrogen. Milk CN consists of four CN fractions (α_{S1}-CN, α_{S2}-CN, β-CN, and κ-CN) in the approximate proportion of 3:1:3:1 (Farrell, 2008) and gamma-casein, which is a product of degradation of β-CN (Ostersen et al., 1997; Miller et al., 1990). Changes in the concentration of individual protein fractions in milk affect various processing attributes of the milk including rennet coagulating time (Ikonen et al., 2004; Joudo et al., 2008), curd firmness (Ikonen et al., 2004; Wedholm et al., 2006; Joudo et al., 2008), pH (Ikonen et al., 2004; Joudo et al., 2008), and cheese yield (Wedholm et al., 2006; Bonfatti et al., 2011a).

Results from numerous studies reveal that parity, stage of lactation (Ostersen et al., 1977; Kroeker et al., 1985; Ng-Kwai-Hang et al., 1987), and breed (Cerbulis et al., 1975; Auldist et al., 2004; Joudu et al., 2008; Lopez-Villalobos, 2012) are associated with the individual protein fractions of milk. For example, mean α_S-CN was reported to increase between first and third parity cows but plateaued thereafter, whereas β-CN decreased as parity number increased (Kroeker et al., 1985; Kwai-Hang et al., 1987). Changes in the proportions of α_{S1}-CN, β-CN, and κ-CN in milk have also been associated with stage of lactation and cow parity (Kroeker et al., 1985).

Free amino acids (FAA) in milk may be used as human nutritional supplements since they are more digestible than protein (Mero, 1999; Gleeson, 2008). For milk processing purposes, a high level of FAA is undesirable, as FAA are a result of deprotenization and an indication of poorer quality milk. Therefore, bovine milk is less suitable for processing in early and late lactation, when total FAA are in greatest concentration (Auldist et al., 1995, **Chapter 2**). Nevertheless, to-date, no study has investigated the variability in FAA across different parities, milking times of the day, calendar months of the year at test or dairy breeds.

Mid-infrared spectroscopy (MIRS) is commonly used worldwide to predict milk fat, protein, casein and lactose of individual animal and bulk tank milk samples. Previous studies propose MIRS as a rapid and cost-effective analytic tool for recording phenotypes at population level (De Marchi et al., 2014; McParland and Berry, 2016). The ability of MIRS to predict milk technological traits, detailed protein composition (α_{S1}-CN, β-CN, κ-CN, α-LA, β-LG A, and β-LG B), FAA, and milk colour characteristics has been previously documented (Visentin et al., 2015; **Chapter 2; Chapter 3**). The objective of the present study was to quantify the associations between cow-level factors, as well as other factors, with detailed protein and FAA composition of bovine milk predicted using MIRS.

4.3 Materials and methods

4.3.1 Spectral Data

Milk samples were available from two sources; (i) seven research herds operated by the Animal and Grassland Research and Innovation Center (Teagasc, Ireland), and (ii) 69 Irish commercial dairy herds. Spectra from the research herds comprised of 126,845 separate morning and evening milk samples from 2,535 lactations and 1,439 cows. Spectra from the commercial herds comprised 44,976 milk samples (morning and evening milk samples combined) from 14,874 lactations and 8,733 cows. Milk samples were form two seasonal calving systems (spring and autumn). Milk chemical composition (milk fat, protein, casein, and lactose concentration) was predicted for all milk samples using the same Fourier transform infrared spectrometer (Foss MilkoScan FT6000 (Foss Electronic A/S, Hillerød, Denmark) based at the Animal and Grassland Research and Innovation Center, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland. The generated spectrum, containing 1,060 transmittance data in the mid-infrared region between 900 and 5,000 cm^{-1} was stored.

Mid-infrared spectroscopy models were developed using partial least squares regressions (Proc PLS; SAS Institute Inc.) with untreated spectra as described in detail in **Chapter 2**. Spectral regions from 926 to 1,580cm^{-1}, from 1,717–2,986cm^{-1}, and from 3,696–3,808 cm^{-1} were used to develop all prediction models based on the observed loadings for each wavelength. In brief, between the

years 2013 and 2014, a calibration dataset was generated using 715 individual milk samples from the same seven research herds used in the present study. Milk samples from 345 cows appeared in both datasets. Spectral outliers were determined as milk samples with a mahalanobis distance greater than three (Williams, 2007) relative to the mean of the calibration dataset. Prediction models were developed using 557 reference values for individual proteins and up to 715 reference values for FAA determined by high performance liquid chromatography. The accuracy (i.e. the coefficient of correlation) of prediction from 1) split-sample cross validation and 2) from external validation on an independent 25% of the data (not included in model calibration), are reported in **Chapter 2.** The accuracy of prediction in external validation was, on average, moderate; the coefficient of determination for protein fractions ranged from 0.39 (β-LG A) to 0.74 (total CN) and for FAA ranged from 0.26 (Arg) to 0.75 (Gly; **Chapter 2**).

4.3.2 Data Editing

Spectral data with a Mahalanobis distance greater than three (Williams, 2007) relative to the mean of the 715 samples used in **Chapter 2** to develop the prediction equations, were considered as spectral outliers (1,634 milk samples in total) and discarded. Furthermore, predicted values of proteins and FAA which were >3 standard deviations from the mean of the reference trait were considered to be outliers and also removed. Only milk yield over a 24 h period was available for the commercial cows; therefore milk yield over a 24 h period was computed for the research cows, as the sum of their morning and evening milk yield from the same day. Only milk samples from recorded between 5 and 305 DIM and from parities ≤10 were retained for analysis; parties greater than 5 were grouped together for analysis.

Contemporary group of experimental treatment by test-date was defined for milk samples from cows in research herds, whereas contemporary group of herd-test-date was defined for milk samples from cows in commercial herds. Only records within contemporary groups with at least ten records were retained for analysis. The research and commercial data sets were combined for analysis. After editing, the final data set comprised of 134,546 milk spectra from 9,572 cows.

Pedigree data and breed composition of all animals were available from the national database managed by the Irish Cattle Breeding Federation (http://www.icbf.com). Only milk samples from Holstein (HO), Friesian (FR), Jersey (JE), Norwegian Red (NR), and Montbelliarde (MO) cows as well as their crosses (HO x FR, HO x JE, HO X NR, HO X MO, JE X FR, JE X NR, JE X MO, NR X FR, and MO X FR) were retained for analysis. The data consisted of 6,724 purebred cows (i.e., ≥75% pure), 2,848 crossbred cows and 1,853 cows with crossbred parents. The number of records, cows, number of lactations, and average parity of each breed and cross are in **Table 4.1**. Coefficients of heterosis and recombination loss were calculated for each cow as:

$$\text{heterosis} = 1 - \sum_{i=1}^{n} sire_i * dam_i$$

and

$$\text{recombination loss} = 1 - \sum_{i=1}^{n} \frac{sire_i^2 + dam_i^2}{2}$$

where $sire_i$ and dam_i are the proportion of genes of the breed i in the sire and the dam, respectively (VanRaden and Sanders, 2003).

4.3.3 Data Analysis

Factors associated with both protein and FAA composition traits were quantified separately using linear mixed models in ASReml (Gilmour et al., 2009). Factors considered in the model included the fixed effects of calendar month of milk test, milking time of the day (AM, PM or both combined), parity (1, 2, 3, 4, 5 and ≥6), stage of lactation (in 30 day intervals), the observed interaction between parity and stage of lactation, breed proportion of the cow fitted as separate covariates (Friesian, Jersey, Norwegian Red, Montbelliarde, and other), and general heterosis and recombination loss coefficients of the cow; Holstein breed proportion was not included in the model to avoid linear dependencies and therefore breed solutions reported are relative to a Holstein cow. The random effects of contemporary group as well as both within and across lactation effects were included in all models. Least square means were estimated based on a reference cow represented as a 100% Holstein, parity 3 cow, milked in the morning, averaged across stages of lactation and calendar months of the year at test. In a separate series of analyses, models with a protein fraction as the dependent variable were adjusted for total milk protein content (i.e., included as a

covariate); models with FAA as the dependent variable were adjusted for 24 h milk yield by including 24 h milk yield as a covariate in the models.

4.4 Results

4.4.1 Descriptive Statistics

Mean predicted values of all milk traits in the research and commercial herds are summarised in **Table 4.1.** Mean values were similar for both the research and commercial herds. Mean values of total CN (α_{S1}-CN, α_{S2}-CN, β-CN, and κ-CN) and total whey (α-LA, β-LG A, and β-LG B) for both research and commercial herds, respectively, were approximately in the ratio 6:1. Individual CN fractions (α_{S1}-CN, α_{S2}-CN, β-CN, and κ-CN) were present in the ratio 4:1:4:2 for both the research and commercial herds, respectively. The coefficient of variation (CV) differed among traits, and ranged from 10% (β-CN) to 51% (β-LG-B) for the protein fractions. The FAA present in the greatest quantity in the milk was Glu which represented 57% of total FAA (Glu, Gly, Lys, Arg, Asp, Ser, and Val) in the milk. Contemporary group accounted for between 45.34% (total FAA) and 86.11% (α-LA) of the variability in the traits investigated.

4.4.2 Non-Genetic Factors

Evening milk had a greater ($P<0.001$) concentration of α_{S1}-CN, α_{S2}-CN, total whey protein, α-LA, total β-LG, and β-LG A, but a reduced ($P<0.01$) concentration of β-CN (13.60g/L vs 13.75 g/L) compared to morning milk when adjusted for crude protein content (**Table 4.3**). Furthermore, although evening milk has more ($P<0.01$) β-CN than morning milk the biological difference is small. Evening milk had a greater ($P<0.001$) concentration of all FAA (total FAA, Glu, Gly, Lys, Arg, Asp, Ser, and Val) compared to morning milk when adjusted for milk yield (**Table 4.3**).

The observed interaction between stage of lactation and parity on the concentration of protein fractions persisted irrespective of whether or not adjustments were made in the statistical model for either differences in crude protein content or 24 h milk yield (results not shown). When adjusted for crude protein content, total CN and protein fractions (except for α-LA) decreased post-calving to between 36-65 DIM across all parties but gradually increased thereafter

(**Figure 4.1, Figure 4.2 and Figure 4.4**). In younger parity cows, α-LA in milk decreased between 5-155 DIM ($P<0.05$) and then plateaued when adjusted for crude protein content; however, in older cows, α-LA in milk remained constant until mid to late lactation, after which it decreased in concentration (**Figure 4.4**). Across all stages of lactation, younger animals produced milk with a greater concentration of total CN ($P<0.05$), total whey ($P<0.001$), and total β-LG ($P<0.001$), than their older contemporaries when adjusted for crude protein content.Younger cows produced more α_{S1}-CN ($P<0.01$) and β-CN ($P<0.001$) in milk than older cows, but first parity cows produced less α_{s2}-CN (0.001) and κ-CN (0.01) than multiparous cows when adjusted for crude protein content.

The interaction between stage of lactation and parity ($P<0.001$) on total FAA, Glu, Gly, and Lys concentration in milk adjusted for milk yield is illustrated in **Figure 4.3**. Irrespective of cow parity, Lys and Val concentration decreased in milk until 36-65 DIM, subsequently, Lys concentration plateaued and Val concentration continued to increase across stage of lactation. Total FAA and Gly concentration decreased from 5-125 DIM after which, total FAA continued to decrease across stage of lactation in earlier parities but plateaued in later parities and Gly concentration plateaued irrespective of parity. Across stage of lactation, younger cows had a greater ($P<0.001$) concentration of total FAA, Glu, and Asp in milk compared to older contemporaries. The concentration of Gly was the same across parities ($P<0.05$) and there was a lower concentration of Lys and Arg in earlier parities compared to later parities ($P<0.001$).

Table 4.1 Number of records, cows and lactation records and average parity for different breeds and crosses used in the present study.

Breed[1]	N	Cows	Lactations	Parity
	≥87.5%			
HO	37,929	4,684	7,410	2.86
FR	8	4	4	3.25
JE	4572	55	96	2.14
NR	48	5	5	6.00
MO	114	52	78	4.18
HO X FR	58,556	3,644	6,006	2.58
HO X JE	25,080	522	912	2.51
HO X NR	2,506	287	530	2.27
HO X MO	1,140	175	274	3.25
JE X FR	2,899	70	121	2.79
JE X NR	1,518	51	102	2.25
JE X MO	51	1	1	1.00
NR X FR	75	12	20	2.29
MO X FR	50	10	17	4.28
	≥75%			
HO	62,653	6,562	10,568	2.70
FR	288	31	48	2.60
JE	4,872	63	107	2.22
NR	179	13	21	4.94
MO	135	55	81	4.46
HO X FR	34,690	1,826	2,954	2.68
HO X JE	24,204	488	851	2.49
HO X NR	2,280	271	501	2.07
HO X MO	728	122	188	3.65
JE X FR	2,823	67	117	2.79
JE X NR	1,518	51	102	2.25
JE X MO	51	1	1	1.00
NR X FR	75	12	20	2.29
MO X FR	50	10	17	4.28

[1]HO = Holstein, FR = Friesian, JE = Jersey, HO×FR = Holstein-Friesian cross, HO×JE = Holstein-Jersey cross, HO×NR = Holstein -Norwegian Red cross, HO×MO = Holstein -Montbelliarde cross, JE×FR = Jersey-Friesian cross, JE×NR = Norwegian Red cross, JE × MO = Jersey-Montbelliarde cross, NR×FR= Norwegian Red-Friesian cross, MO×FR= Montbelliarde-Friesian cross; a purebred animal was deemed to be ≥87.5% of the breed or ≥75% of the breed.

Table 4.2 Number of records (N), mean (SD in parentheses), and coefficient of variation (CV) of the studied traits predicted using mid-infrared spectroscopy in research and commercial herds.

	Research herds			Commercial Herds			
Trait	N	Mean (SD)	CV, %	N	Mean (SD)	CV, %	CV of Contemporary Group, %
Protein g/L milk							
Total CN	94,148	36.37 (4.48)	12.31	40,247	35.66 (3.95)	11.06	56.46
α_{S1}-CN	94,160	13.84 (1.89)	13.67	40,249	13.38 (1.68)	12.57	58.95
α_{S2}-CN	94,235	3.68 (0.50)	13.63	40,247	3.62 (0.46)	12.75	53.40
B-CN	94,080	12.94 (1.55)	11.97	40,155	12.99 (1.30)	10.05	62.34
K-CN	94,211	6.08 (0.96)	15.76	40,252	5.90 (0.86)	14.62	62.87
Total Whey	94,147	6.16 (1.59)	25.78	40,236	6.07 (1.42)	23.38	66.51
α-LA	94,185	1.09 (0.22)	19.73	39,921	1.15 (0.16)	13.87	86.11
Total β-LG	94,072	5.13 (1.57)	30.64	40,164	4.98 (1.42)	28.57	67.63
β-LG A	94,231	2.45(0.52)	21.35	40,258	2.27 (0.49)	21.43	75.20
β-LG B	89,467	2.46 (1.26)	51.14	38,924	2.95 (1.28)	43.45	52.71
Free AA μg/mL milk							
Total free AA	94,286	53.70 (18.26)	34.00	40,260	53.38 (16.43)	30.78	45.34
Glu	94,286	30.64 (13.89)	45.33	40,260	31.50 (11.27)	35.78	45.46
Gly	94,286	7.94 (5.47)	68.92	40,260	8.27 (5.11)	61.82	51.18
Lys	94,286	4.82 (3.11)	64.52	40,260	4.55 (2.71)	59.46	71.09
Arg	94,286	3.39 (1.20)	35.41	40,260	3.36 (1.46)	43.42	64.52
Asp	94,286	2.75 (1.47)	53.36	40,260	2.70 (1.51)	55.93	63.88
Ser	94,286	2.60 (1.83)	67.94	40,260	1.40 (0.65)	46.32	47.47
Val	94,286	1.48 (0.79)	53.19	40,260	1.59 (0.86)	54.25	61.42

Table 4.3 Least square means (SE in parentheses) of individual protein fractions adjusted for crude protein content and of individual FAA (µg/ml milk) adjusted for milk yield in both morning (AM) and evening (PM) milking[1].

	Milking time		
	AM	PM	
Protein g/L milk			
Total CN	38.43 (0.0979)	38.66 (0.0983)	
αs_1-CN	14.39 (0.0414)	14.80 (0.0415)	P <0.001
αs_2-CN	3.91 (0.0105)	3.99 (0.0106)	P <0.001
β-CN	13.72 (0.0323)	13.60 (0.0324)	P <0.01
κ-CN	6.55 (0.0201)	6.58 (0.0202)	
Total Whey	6.41 (0.0372)	6.68 (0.0374)	P <0.001
α-LA	1.13 (0.0046)	1.16 (0.0047)	P <0.001
Total β-LG	5.33 (0.0359)	5.61 (0.0360)	P <0.001
β-LG A	2.49 (0.0125)	2.61 (0.0125)	P <0.001
β-LG B	2.60 (0.0360)	2.70 (0.0361)	P < 0.05
Free AA µg/mL milk			
Total FAA	53.73 (0.3871)	60.73 (0.3890)	P <0.001
Glu	30.01 (0.3138)	34.09 (0.3152)	P <0.001
Gly	7.84 (0.0878)	8.64 (0.0886)	P <0.001
Lys	5.08 (0.0646)	5.87 (0.0651)	P <0.001
Arg	3.53 (0.0272)	4.18 (0.0274)	P <0.001
Asp	2.74 (0.0313)	3.17 (0.0314)	P <0.001
Ser	2.93 (0.0315)	2.88 (0.0318)	
Val	1.60 (0.0185)	1.79 (0.0187)	P <0.001

[1]Commercial herds are not included

After adjusting for crude protein content a peak in the concentration of all CN fractions was evident in the months of August, September and October (**Figure 4.5**; P<0.001); while the concentration of α-LA remained relatively constant across the year (P<0.05). The concentration of Glu was greater (P<0.001) during the months of February, March, April and June, while the concentration of Gly was greater (P<0.001) during the months of February, March and June when adjusted for milk yield. The change in the concentration of Asp, Ser and Val across calendar month of the year adjusted for milk yield was small but different to zero (P<0.05).

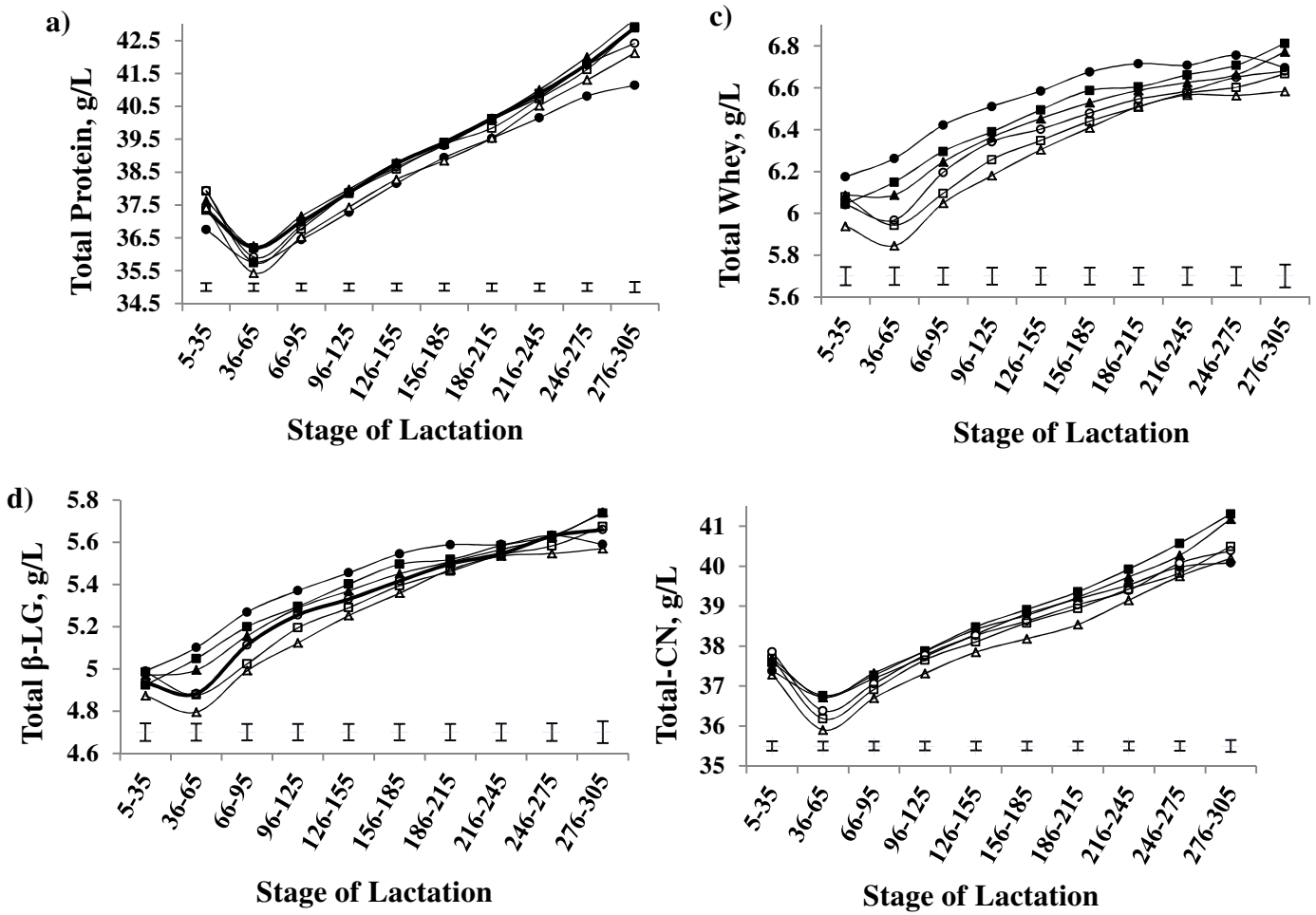


Figure 4.1 Trends in concentration of proteins in milk a) total protein, b) total CN, c) total whey and d) total β-LG adjusted for crude protein content across stage of lactation for parity 1 (●), parity 2 (■), parity 3 (▲), parity 4 (○), parity 5 (□) and parity ≥ 6 (Δ) animals. Error bars represent the mean SE across parities.

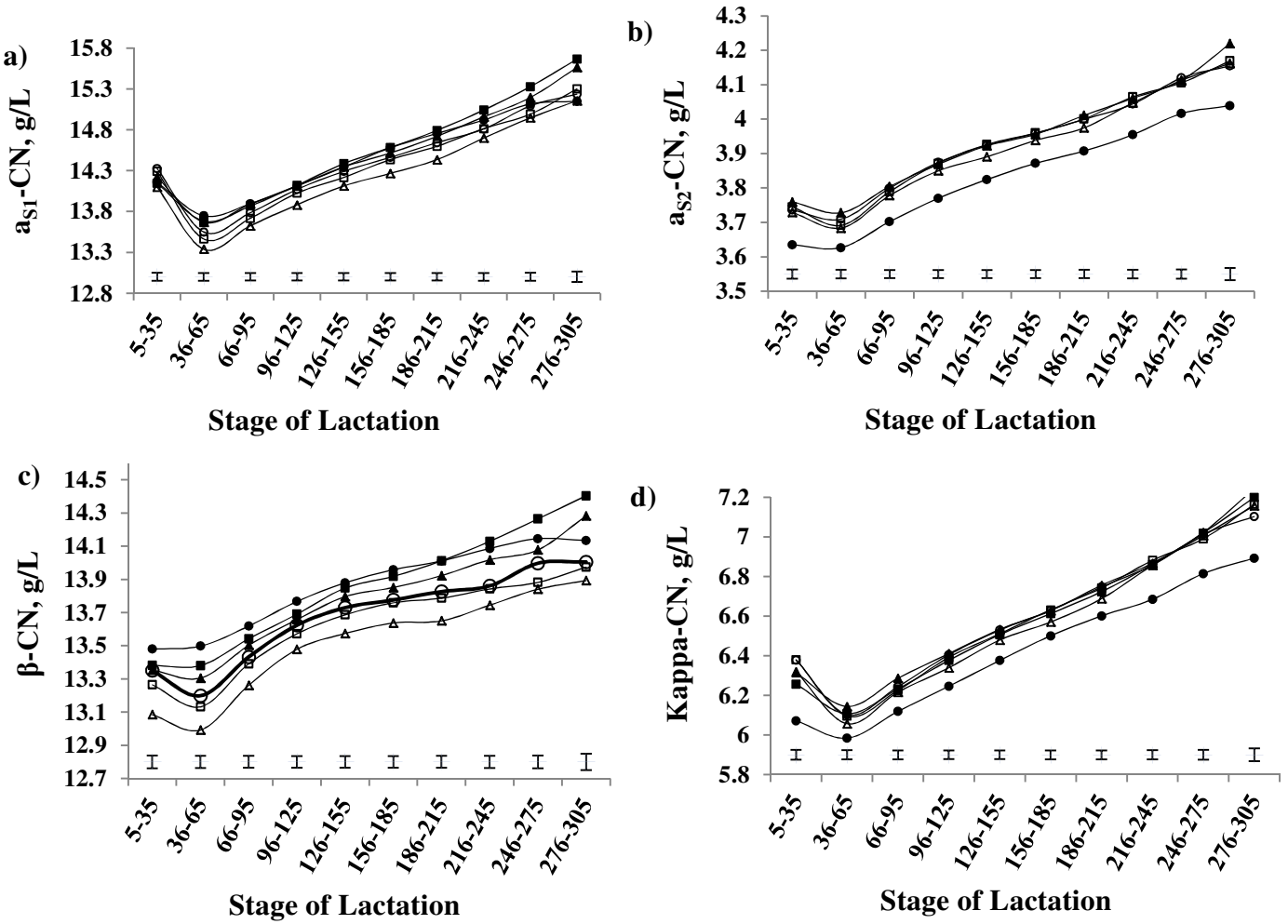
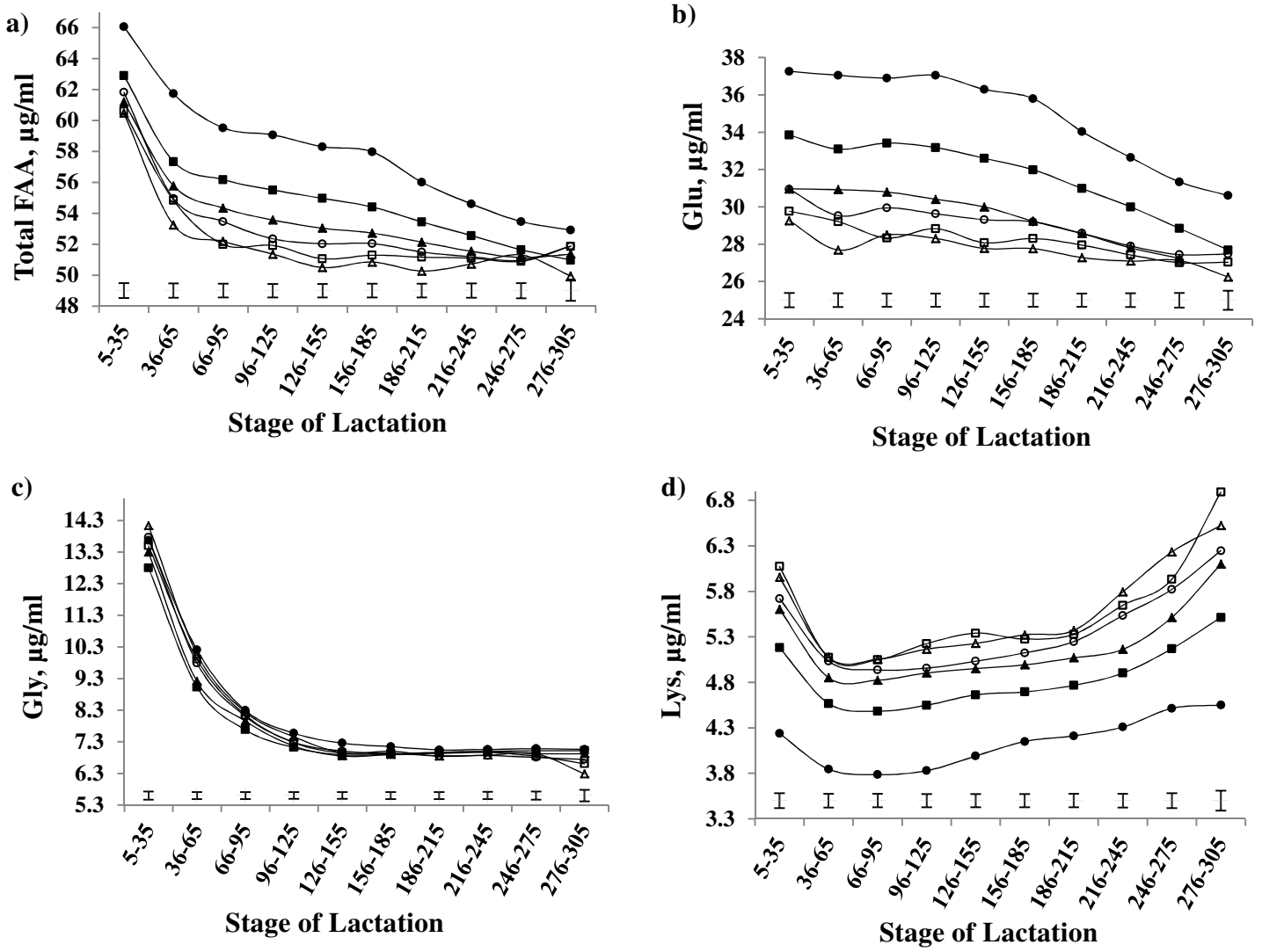


Figure 4.2 Trends in concentrations of casein fractions in milk a) αs_1-CN, b) αs_2-CN, c) β-CN and d) κ-CN adjusted for crude protein content across stage of lactation for parity 1 (●), parity 2 (■), parity 3 (▲), parity 4 (○), parity 5 (□) and parity ≥ 6 (Δ). Error bars represent the mean SE across parities.


a)
Total FAA, µg/ml
66
64
62
60
58
56
54
52
50
48
5-35
36-65
66-95
96-125
126-155
156-185
186-215
216-245
246-275
276-305
Stage of Lactation
b)
Glu, µg/ml
38
36
34
32
30
28
26
24
Stage of Lactation
c)
Gly, µg/ml
14.3
13.3
12.3
11.3
10.3
9.3
8.3
7.3
6.3
5.3
Stage of Lactation
d)
Lys, µg/ml
6.8
6.3
5.8
5.3
4.8
4.3
3.8
3.3
Stage of Lactation

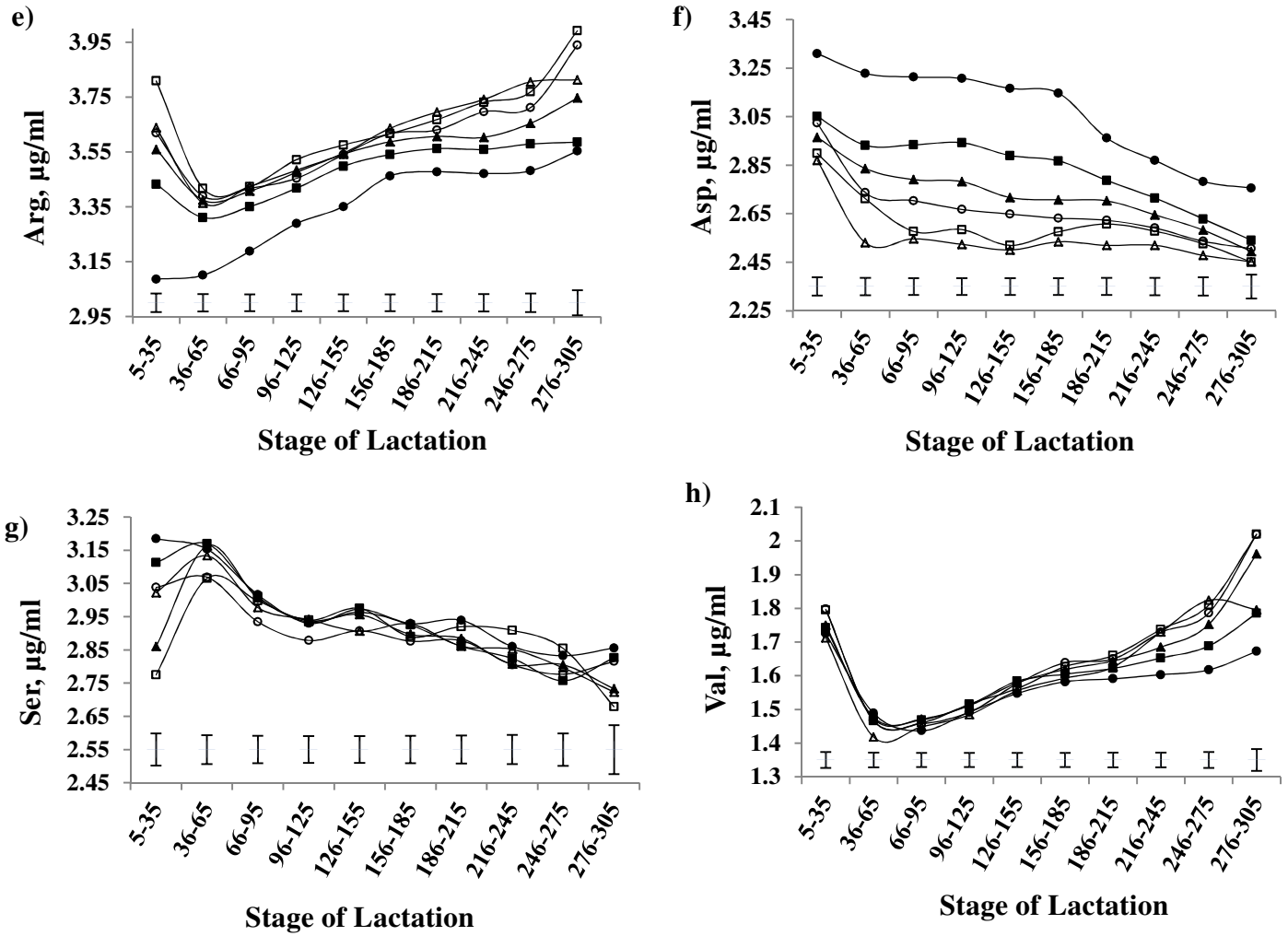


Figure 4.3 Trends in total and individual FAA in milk a) Total FAA, b) Glu, c) Gly, d) Lys, e) Arg, f) Asp, g) Ser and h) Val adjusted for milk yield across stage of lactation for parity 1 (●), parity 2 (■), parity 3(▲), parity 4 (○), parity 5 (□) and parity ≥ 6 (Δ) animals. Error bars represent the mean SE across parities.

4.4.3 Breed, Heterosis and Recombination effects

Breed regression coefficient estimates for concentration of proteins adjusted for crude protein content expressed relative to a purebred Holstein for Friesian, Jersey, Norwegian Red and Montbelliarde breeds and associated heterosis and recombination estimates are provided in **Table 4.4.** Jersey cows produced milk with the greatest concentration of all CN fractions (P<0.001), and produced milk with 3.91 g/L, 3.14 g/L, 2.86 g/L, and 4.64 g/L more total CN than Holstein, Friesian, Norwegian Red, and Montbelliarde cows, respectively. Also, Jersey cows produced milk that had a greater concentration of the casein fractions (α_{S1}-, α_{S2}-, β-, and κ-CN) in addition to a greater concentration total whey, α-LA, total β-LG, and β LG A relative to Holstein cows. The concentration of total whey protein in milk of Jersey cows was 0.42 g/L, 0.23 g/L, 0.42 g/L, and 0.53 g/L greater when compared to the milk of Holstein, Friesian, Norwegian Red and Montbelliarde cows, respectively. Jersey cows produced less (P<0.001) total FAA, Glu, Gly and Asp than any other breed of cow, including Holsteins (**Table 4.5**).

Both heterosis and recombination estimates for all traits were small in magnitude. Relative to the purebred parent average, first-cross (F_1) cows produced milk that had 0.27g/L more total CN and 0.13 g/L more β-CN. Positive recombination estimates were observed for total CN, α_{S1}-CN, β-CN, κ-CN, total β-LG, and β-LG A. in milk. Heterosis estimates for all FAA (except Lys) in milk were not different from zero and recombination estimates for all FAA were not different from zero.

Table 4.4 Breed regression coefficient estimates (SE in parentheses) for concentration of proteins (g/L milk) adjusted for crude protein content expressed relative to a purebred Holstein for Friesian (FR), Jersey (JE), Norwegian Red (NR) and Montbelliarde (MO), and associated heterosis and recombination estimates.

	HO	FR	JE	NR	MO	Heterosis	Recombination
Total CN	0^{a}	0.77^{b} (0.21)	3.91^{c} (0.19)	1.05^{d} (0.29)	-0.73^{ae} (0.32)	0.27 (0.10) *	1.05 (0.16) *
α_{S1}-CN	0^{a}	0.32^{b} (0.09)	1.72^{c} (0.08)	0.49^{b} (0.12)	-0.38^{ad} (0.14)	0.10 (0.05) *	0.50 (0.07) *
α_{S2}-CN	0^{a}	0.06^{ab} (0.02)	0.42^{c} (0.02)	0.06^{b} (0.03)	-0.08^{d} (0.03)	0.02 (0.01) *	0.07 (0.02)*
β-CN	0^{a}	0.27^{b} (0.06)	0.98^{c} (0.06)	0.23^{ab} (0.09)	-0.22^{ad} (0.10)	0.13 (0.03) *	0.27 (0.05) *
κ-CN	0^{a}	0.07^{ab} (0.04)	0.69^{c} (0.04)	0.19^{b} (0.05)	-0.08^{ad} (0.06)	0.03 (0.02)	0.14 (0.03)*
Total Whey	0^{a}	0.19^{ab} (0.07)	0.42^{c} (0.07)	0.00^{abd} (0.10)	-0.11^{ad} (0.11)	-0.04 (0.04)	0.22 (0.06)*
α-LA	0^{a}	0.03^{ab} (0.01)	0.08^{c} (0.01)	0.01^{ad} (0.01)	-0.01^{ad} (0.01)	0.01 (0.01)	0.03 (0.00)*
Total β-LG	0^{a}	0.12^{ab} (0.07)	0.34^{c} (0.07)	0.00^{abd} (0.10)	-0.11^{abd} (0.11)	-0.04 (0.04)	0.22 (0.06)*
β-LG A	0^{a}	$0.07^{a\,b}$ (0.02)	0.32^{c} (0.02)	0.07^{ab} (0.03)	-0.12^{d} (0.03)	0.02 (0.01)*	0.10 (0.02)*
β-LG B	0^{a}	0.11^{abcde} (0.08)	-0.04^{abcde} (0.07)	-0.09^{abcde} (0.11)	0.09^{abcde} (0.12)	-0.09 (0.04)*	0.13 (0.06)*

[a-e]Values within rows differing in superscript are different ($P < 0.05$)

*Regression coefficients significantly different to zero ($P < 0.05$)

Table 4.5 Breed regression coefficient estimates (SE in parentheses) for concentration of free AA (µg/ml milk) adjusted for milk yield expressed relative to a purebred Holstein for Friesian (FR), Jersey (JE), Norwegian Red (NR) and Montbelliarde (MO), and associated heterosis and recombination estimates.

	HO	FR	JE	NR	MO	Heterosis	Recombination
Free AA, µg/ml							
Total free AA	0^{a}	3.09^{b} (0.86)	-7.35^{c} (0.76)	4.42^{bd} (1.20)	2.28^{abd} (1.35)	-0.06 (0.43)	-0.66 (0.68)
Glu	0^{a}	2.63^{b} (0.70)	-6.97^{c} (0.62)	3.39^{bd} (0.98)	1.89^{abd} (1.09)	-0.17 (0.35)	-1.08 (0.55)
Gly	0^{a}	0.46^{b} (0.16)	-0.52^{c} (0.14)	0.36^{abd} (0.22)	0.74^{bd} (0.29)	-0.01 (0.08)	0.21 (0.12)
Lys	0^{a}	-0.27^{ab} (0.10)	0.59^{c} (0.10)	-0.07^{abd} (0.14)	-0.43^{bd} (0.17)	0.13 (0.05) *	0.06 (0.08)
Arg	0^{a}	-0.13^{ab} (0.05)	0.02^{ac} (0.05)	0.04^{acd} (0.07)	-0.07^{abcd} (0.08)	-0.01 (0.03)	-0.04 (0.04)
Asp	0^{a}	0.11^{ab} (0.06)	-0.56^{c} (0.06)	0.29^{b} (0.09)	-0.02^{ab} (0.10)	0.02 (0.03)	0.01 (0.05)
Ser	0^{a}	0.05^{abcde} (0.05)	-0.01^{abcde} (0.04)	0.00^{abcde} (0.07)	0.06^{abcde} (0.10)	0.02 (0.03)	0.02 (0.04)
Val	0^{a}	0.05^{ab} (0.03)	-0.15^{c} (0.03)	0.22^{d} (0.05)	-0.01^{ab} (0.06)	0.00 (0.02)	0.00 (0.03)

$^{a-e}$Values within rows differing in superscript are different ($P < 0.05$)

* Regression coefficients significantly different to zero ($P < 0.05$)

4.5 Discussion

The objective of the present study was to determine the factors associated with protein composition and FAA composition of bovine milk. In the present study, mean values of α_{S1}-, α_{S2}-, β-, and κ-CN were in the ratio 4:1:4:2, which was not in agreement with results by Farrell et al. (2004), who documented a ratio of 3:1:3:1. However the ratio of total CN to total whey protein (6:1) in the present study was consistent with Farrell et al. (2004). The CV for total whey protein and whey fractions in the present study was greater than that reported by De Marchi et al. (2010) for 1,336 Simmental cows only; the present study, however, contained records from cows of multiple breeds and crossbreds. Total mean proteins, determined using HPLC (42.53 g/L for the research herds; 41.73 g/L for the commercial herds), were higher than that recorded using MIRS (37.45 g/L for the research herds; 36.67 g/L for the commercial herds). This is most likely due to cumulative variation during summation of the individual protein values when integrating of peak areas from the HPLC data (**Chapter 2**). De Marchi et al., (2010) and Bonfatti et al., (2011a) also reported high mean protein values using HPLC, i.e., up to 40.12 g/L and 40.68 g/L respectively, in milk from Simmental cows.

Similar trends across stage of lactation for α_{S1}-CN, β-CN, β-LG, and α-LA in milk adjusted for crude protein content (i.e. decrease in early lactation followed by a gradual increase) were observed both in the present study and elsewhere (Ng-Kwai-Hang et al., 1987). Kroeker et al. (1985) observed a similar stage of lactation trend when β-CN was expressed relative to total casein. The observed decline in total proteins and protein fractions in early lactation coincides with the period of negative energy balance typically seen in dairy cows in early lactation (Berry et al., 2006); negative protein balance may also occur (when the amount of protein broken down by the cow exceeds the amount ingested by the cow). It can take a cow up to 20 weeks to regain a positive energy and protein balance, and for actual milk protein content to increase again (Taylor et al., 2003).

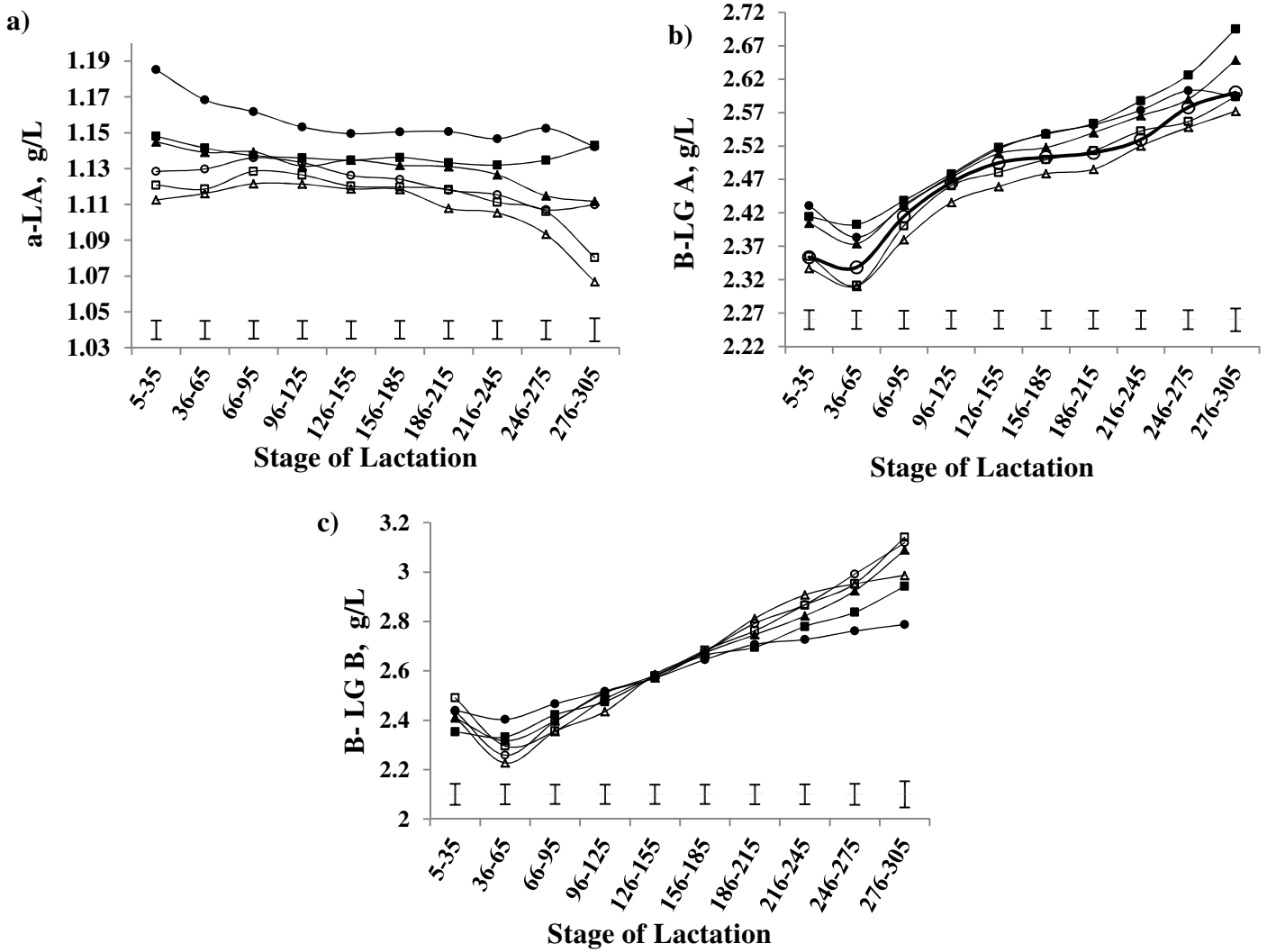


Figure 4.4 Trends in concentrations of whey proteins in milk a) α-LA, b) β-LG A and c) β-LG B adjusted crude protein content across stage of lactation for parity 1 (●), parity 2 (■), parity 3 (▲), parity 4 (○), parity 5 (□) and parity ≥ 6 (Δ). Error bars represent the mean SE across parities.

The reduction in concentration of total FAA and the individual FAA (Glu, Gly, Lys, Arg, and Asp) in milk up to 65 days post-calving corroborates documented reports by Ghadimi and Pecora (1963), who studied the concentration of these FAA in bovine milk from 7 to 60 DIM. Both total FAA and Glu in milk decreased as parity increased and, despite the part-whole relationship between them (Glu makes up over 55% of total FAA), the lactation profile of Glu and total FAA in milk differed in younger animals (**Figure 4.3**). To our knowledge, no study to date has investigated the association between parity and FAA composition of milk.

4.5.1 Variability in milk quality

Considerable variability in protein fractions and FAA existed in both populations studied; the CV of β-LG B and Gly were 51% and 69% in the research herds, respectively, which is considerably greater than the CV of 3.1% observed for milk yield in the same population (results not shown). Although no heritability estimates were generated in the present study, previously reported heritability estimates in dairy cattle range between 0.02 (α_{S1}-CN) and 0.66 (α-LA) for protein fractions (Kreoker et al., 1985; Huang et al., 2012) and to our knowledge no study to date has estimated heritability for FAA. The success of breeding programs for increased milk yield in dairy cows is well recognized (Berry et al., 2014; Berry, 2008; Norman and Powell, 1999); therefore based on the heritability estimates that exist in the literature for protein fractions, it is possible to assume that breeding for improved protein fractions in milk is possible. The potential of milk MIRS to predict protein fractions and FAA (**Chapter 2**) provides an opportunity to generate large quantities of data for use in genetic evaluations and thus breeding programs. The attributes and tools therefore appear to exist to facilitate breeding programs for superior milk quality characteristics.

Contemporary group accounted for between 45.34% (total FAA) and 86.11% (α-LA) of the variation in the dataset used in this study, indicating that the combination of herd and test-date have large effects on both protein and FAA composition of the milk. These differences offer the potential for herds to be selected on the basis of their protein and FAA profile; further milk price premiums could be paid to herds that are producing milk with a protein or FAA profile that better fits the processors needs for production. Herd-level estimates of milk quality can be readily obtained as a by-product from national genetic evaluations and thus the data can be readily available; these herd solutions would be independent of genetic merit of the producing animals and therefore more closely

reflect the management influence on milk quality. Moreover, the ability to monitor the trend in milk quality over time within a herd will enable the provision of decision support information to producers and processors on the factors affecting the quality of their milk.

4.5.2 Decision support tool

The observed trends in milk protein fractions and FAA across month of the year (**Figure 4.5**) suggest differences in the suitability of milk to produce milk products across the year and provides evidence of the difficulty in acquiring a stable product of constant composition across time. The impact on consistency of product is further compounded in seasonal calving herds as exist in Ireland (Berry et al., 2013) and elsewhere since milk protein fractions and FAA also vary across stage of lactation which is synchronized with calendar month. The structure of the data, coupled with the statistical model, implies that the observed effects reported in the present study are independent of each other and are therefore additive. Commercially available infant formulas have a ratio of total casein to total whey protein close to that of human milk, but, β-LG which is not present in human milk, is present in the greatest amount in cow milk. It is advantageous for infant formula producers to select cow milk with a higher concentration of α-LA and a lower concentration of β-LG; **Figure 4.4** demonstrates that younger parity animals in early lactation produce the highest concentration of α-LA. Results from the present study also show that milk produced by young Jersey cows in the months of August, September and October could achieve a greater concentration of CN fractions in milk.

Bovine and human milk also differ in their amino acid profile (Chuang et al., 2005); for example human milk has, on average, four times more Arg and twice as much Tau as bovine milk (Sarwar et al., 1998). Free amino acids are often added to infant formula by processors, especially for the production of formula for infants with allergies to casein and whey protein fractions in milk (Owens et al., 2013). Results from the present study may also aid infant formula processors to select milk naturally higher in the sought after protein or FAA profile, thereby minimizing the requirement for protein and FAA additives.

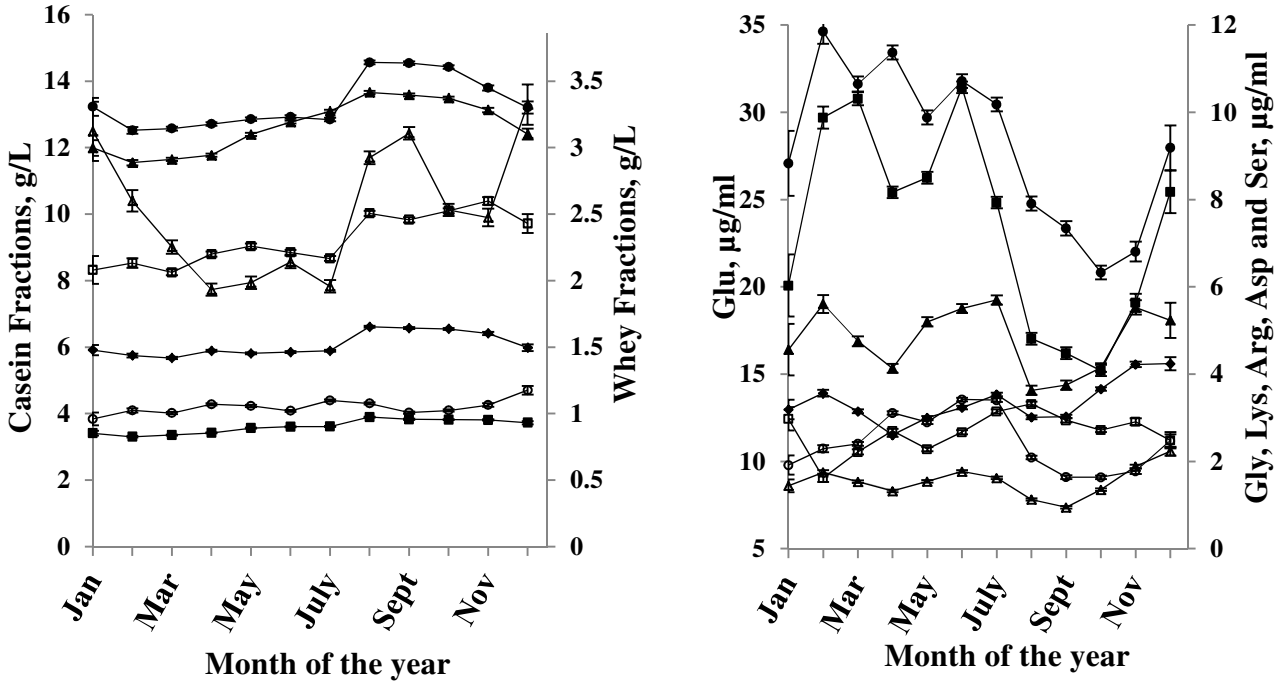


Figure 4.5 a) Least square means of concentrations (g/L milk) of casein fractions [αs1-CN (-●-; primary vertical axis), αs2-CN (-■-; primary vertical axis), β-CN (-▲- ; primary vertical axis) and κ-CN (-♦-; primary vertical axis)] and whey fractions [α-Lac (-○-; secondary vertical axis), β-LG A (-□-; secondary vertical axis) and β-LG B (-Δ -; secondary vertical axis)] adjusted for total milk protein content across calendar month of the year.

b) Least square means of concentrations (µg/ml milk) of Glu (-●-;primary vertical axis) Gly (-■-; secondary vertical axis), Lys (-▲- ; secondary vertical axis), Arg (-♦-; secondary vertical axis), Asp (-○-; secondary vertical axis), Ser (-□-; secondary vertical axis), Val (-Δ -; secondary vertical axis) in milk adjusted for milk yield across calendar month of the year.

Previous studies have revealed that greater concentrations of all CN fractions in milk significantly increase cheese yield (Wedholm et al., 2006), and rennet coagulating time is positively correlated with the contents and proportions of α_{S1}-CN and α_{S2}-CN in total CN (Bonfatti et al., 2011a). Figures 2 and 3 indicate that the concentration of certain protein fractions (α_{S1}-CN, β-CN, κ-CN, and β-LG B) could possibly be too low for cheese production in early lactation; this could be useful information for cheese manufacturers to optimize quality and yield of cheese. Mid and late lactation milk has a greater concentration of κ-CN than early lactation and parity one animals have a lower concentration of κ-CN than their older contemporaries. A higher concentration of κ-CN in milk results in a smaller casein micelle (Gutierrez-Adan et al., 1996) and therefore in a shorter rennet coagulation time contributing to a stronger curd and more cheese yield. However, it is the genetic polymorphism of κ-CN B which is of importance (Ikonen et al., 1999), suggesting a way to improve milk process ability for cheese production would be selecting animals with genes coding for κ-CN B. Jersey cows produced more casein fractions but less total FAA in milk than Holstein cows, indicating Jersey cows may produce milk more suitable for cheese production and of a better processing quality than Holstein cows.

4.5.3 Crossbreeding

Heterosis is defined as the difference between the performance of a crossbred animal and the average of the parents (Willham and Pollak, 1985). Relative to the purebred parent average, first-cross (F_1) cows produced 0.13 g/L more β-CN. Studies have indicated that the consumption of β-CN A1 is associated with higher mortality rates from coronary heart disease in humans (Laugesen and Elliott, 2003; McLachlan, 2001). However, further research is required to determine the genetic composition of β-CN in the milk analysed in the present study based on its genetic composition. Recombination loss is defined as the disintegration of epistatic associations to form nonparent inter-loci combinations of alleles in crossbred animals (Cassady et al., 2002). Generally recombination has unfavourable effects on milk compositional traits such as protein (Dechow et al., 2007), even though favourable recombination estimates were calculated for concentrations of all the individual protein fractions (total CN, α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, total β-LG, β-LG A and β-LG B) in milk. An unfavourable effect is expected as recombination normally affects traits such as milk production that have been under long-term selection intensity (Sørensen et al., 2008). However, a study by Coffey et al. (2016)

showed recombination to have positive effects on milk compositional traits including protein percentage and suggested that different population breeding goals may be a causative factor towards the inconsistences among studies on the effect of recombination on milk compositional traits. Traditionally, Irish dairy cows may have been naturally selected for fertility and survivability as a result of the seasonal calving system operated in Ireland (Berry et al., 2013), therefore reducing the selection pressure on milk composition traits. Although there are known advantages of crossbreeding (Buckley et al., 2014; Coffey et al., 2016), crossbreeding of dairy cattle is not commonly practised worldwide (Buckley et al., 2014). Results in the present study indicate that crossbred cows had a greater concentration of β-CN in milk than purebred Holsteins, which is advantageous for cheese production demonstrating another advantage to crossbreeding.

4.6 Conclusions

Results from the present study indicate that factors including stage of lactation, parity, calendar month of the year, milking time and breed are all associated with protein and FAA composition of bovine milk. Of particular interest was that younger animals produced more total CN, total whey and total β-LG across early and mid-lactation and more Glu and Asp in milk across lactation than their older contemporaries. Jersey cows produced milk that had a greater concentration of all CN fractions but a lower concentration of total FAA than Holstein cows. Knowledge provided by this study of how individual milk proteins and FAA change across calendar months of the year and across stage of lactation could provide useful input parameters for decision support tools in the management of product portfolios by processors over time.

4.7 Acknowledgements

Funding for this work was received from the Irish Department of Agriculture, Food and the Marine, Research Stimulus Fund project 11/SF/311, Breed Quality.

CHAPTER 5

Genetic and non-genetic factors associated with milk colour in dairy cows

S. Scarso[*], S. McParland[1,†], G. Visentin[*,†], D. P. Berry[†], A. McDermott[†], M. De Marchi[*]

[*]Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro (PD), Italy

[†]Animal and Grassland research and Innovation Center, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland

Accepted for publication in the Journal of Dairy Science

5.1 Abstract

Milk colour is one of the sensory properties that can influence consumer choice of one product over another and it influences the quality of processed dairy products. This study aims to quantify the cow-level genetic and non-genetic factors associated with bovine milk colour traits. A total of 136,807 spectra from Irish commercial and research herds (with multiple breeds and crosses) were used. Milk lightness (L̂*), red-green index (â*), and yellow-blue index (b̂*) were predicted for individual milk samples using only the mid-infrared spectrum of the milk sample. Factors associated with milk colour were breed, stage of lactation, parity, milking-time, udder health status, pasture grazing and seasonal calving. (Co)variance components for L̂*, â*, and b̂* were estimated using random regressions on the additive genetic and within-lactation permanent environmental effects. Greater b̂* value (i.e., more yellow colour) was evident in milk from Jersey cows. Milk L̂* increased consistently with stage of lactation, while â* increased until mid-lactation to subsequently plateau. Milk b̂* deteriorated until 31 to 60 DIM, but then improved until the end of lactation. Relative to multiparous cows, milk yielded by primiparous cows was, on average, lighter (i.e., greater L̂*), more reddish (i.e. greater â*), and less yellow (i.e. lower b̂*). Milk from the morning milk session had lower L̂*, â*, and b̂*. Across the calendar year, L̂* (with the exception of a dip in August) and b̂* generally increased, while â* was relatively constant except for a peak in August. Heritability estimates varied between 0.15 ± 0.02 (30 DIM) and 0.46 ± 0.02 (210 DIM) for L̂*, between 0.09 ± 0.01 (30 DIM) and 0.15 ± 0.02 (305 DIM) for â*, and between 0.18 ± 0.02 (21 DIM) and 0.56 ± 0.03 (305 DIM) for b̂*. For all the three milk colour features, the within trait genetic correlations approached unity as the time intervals compared shortened and were generally < 0.40 between the peripheries of the lactation. Strong positive genetic correlations existed between b̂* value and milk fat concentration, ranging from 0.82 ± 0.19 at 5 DIM to 0.96 ± 0.01 at 305 DIM and confirming the observed phenotypic correlation (0.64, SE=0.01). Results of the present study suggest that breeding strategies for the enhancement of milk colour traits could be implemented for dairy cattle populations. Such strategies, coupled with the knowledge of milk colour traits variation due to non-genetic factors, may represent a tool for the dairy processors to reduce, if not eliminate, the use of artificial pigments during milk manufacturing.

5.2 Introduction

Food colour is known to impact food choice (Fergus, 1993). The sensory properties of milk (i.e., appearance, colour, flavour, aroma, and texture) are also important because of their close relationship with both product quality (Wadhwani and McMahon, 2012) and consumer acceptance (Phillips et al., 1995). The yellow colour of butter and many cheeses is influenced by milk fat carotenoid content (Descalzo et al., 2012), and market preferences for milk fat colour varies across the world (Berry et al., 2009). For example, the yellow colour of dairy products is sometimes said to be associated with a more "green image" by consumers, because of its association with grazing animals (Descalzo et al., 2012). In direct contrast, however, in New Zealand the yellow colour of milk and its associated products are considered an unfavorable attribute in many consumers' opinion (Morris et al., 2002).

Milk colour is known to be affected by many factors including animal genetic merit and breed (Winkelman et al., 1999; Noziere et al., 2006; Berry et al., 2009), stage of lactation and parity (Calderon et al., 2007; Jadhav et al., 2008), time of milking (Quist et al., 2008), udder health status (Espada and Vijverberg, 2002), as well as herd-level factors such as pasture grazing and seasonal calving (Agabriel et al., 2007; Solah et al., 2007; Walker et al., 2013).

To our knowledge no study has attempted to quantify the contribution of genetics to variability in milk colour in terms of L*, a* and b* values. Winkelman et al. (1999) estimated genetic and phenotypic correlations of milk colour traits (in terms of milk colour, fat colour and β-carotene yield) with each other and with milk production traits (milk, fat, and protein yields). Milk colour, in this case, was determined by extraction from milk of the nonsaponifiable compounds. As several studies that investigated food colour used the CIE-L*a*b* method as colour measurement, especially on meat colour (Fletcher, 1999, on broiler meat; Liu et al., 2003, on beef; Zhang et al., 2007, on pork meat), in the present study this method was used to investigate milk colour.

Recently mid-infrared spectroscopy (MIRS) has been demonstrated to be a useful low cost and rapid screening tool (De Marchi et al., 2014) to acquire and predict innovative milk technological phenotypes (Visentin et al., 2015) and determine the b* colour value of milk (**Chapter 3**). Prediction equations developed using MIRS can be used to quantify the milk colour of individual animal samples during routine milk recording as well as more frequently available bulk tank milk samples. Therefore, MIRS

is useful to collate large numbers of unbiased records of milk colour throughout lactation which are estimate animal breeding values.

Thus, the objective of the present study was to quantify the contribution of cow-level genetic and non-genetic factors to variability in milk colour as described by L*, a* and b* indices predicted using MIRS equations.

5.3 Materials and methods

5.3.1 Milk sample collection

A total of 174,062 milk samples were collected between January 2013 and December 2015 from 10,394 dairy cows of five different breeds (Holstein, Friesian, Jersey, Montbeliarde, and Norwegian Red) and crosses. Of these, 129,086 samples were from 1,661 research cows from 7 research farms operated by the Teagasc Animal and Grassland Research and Innovation Centre (Moorepark, Fermoy, Co. Cork, Ireland). Cows in the research herds participated in a series of experimental treatments based on different feeding strategies, stocking rates, calving periods, and length of grazing period. A small proportion of dairy cows in the research herds (90 individuals per year) belonged to the top 1% genetic merit, as ranked based on the national selection index. The remaining 44,976 samples were collected from 8,733 cows from 69 different commercial Irish farms located in South-West Ireland. Cows in research and commercial herds were fed a basal grazed pasture diet, but at times cows in the research farms were supplemented with a small quantity of concentrates (depending on the experimental treatment). All cows were milked twice daily and sampled based on test-day recording system. The average monthly test day records per cow and lactation were 17 and 10, respectively. Coefficients of heterosis and recombination loss were calculated for each cow as $\text{heterosis} = 1 - \sum_{i=1}^{n} sire_i * dam_i$, and $\text{recombination loss} = 1 - \sum_{i=1}^{n} \frac{sire_i^2 + dam_i^2}{2}$, where $sire_i$ and dam_i are the proportion of genes of the breed i in the sire and the dam, respectively (VanRaden and Sanders, 2003). The pedigree of all animals was traced back at least four generations, and comprised a total of 41,232 animals.

For the research data, milk samples were separately collected on consecutive PM and AM milkings once weekly. For commercial herds, a single milk sample was taken during the milk recording day and these samples were collected occasionally (approximately 1,249 spectra/month) and sent for analysis as part of a related research study. Once collected, all samples were analysed within 24 hours (for research samples)

or five days (for commercial samples) in the laboratory of Teagasc Animal and Grassland Research and Innovation Center (Moorepark, Fermoy, Co. Cork, Ireland). Milk chemical composition (protein, fat, lactose, urea, casein, total solids) was predicted using a MilkoScan FT6000 (Foss Electronic A/S, Hillerød, Denmark) and mid-infrared spectra (wavelengths from 900 to 5,000 cm^{-1}) were stored. Somatic cell count (SCC) was determined by Fossomatic (Foss Electronic A/S, Hillerød, Denmark), and converted to somatic cell score (SCS) by taking the $\log_{10}$ of SCC.

5.3.2 Gold standard analysis and prediction model development

Milk colour was measured on a selection of samples for the development of MIR prediction equations using a Chroma Meter CR400 (Konica Minolts Sensing Europe, Nieuweigein, the Netherlands, with viewing geometry d/0) with a closed cone, set on L*, a* and b* system. The selection of samples was discussed in detail in **Chapter 3**. The Chroma Meter CR was calibrated on a white tile. Sub-samples of 10-mL were measured in a cuvette and results were expressed in CIE-L*a*b* uniform colour space. This method is a three-dimensional opponent colour system that represents lightness (L*), red-green (a*) and yellow-blue (b*) values on three axes. The central vertical axis represents the L* index, whose values run from 0 (black) to 100 (white). On each axis the values run from positive to negative. Positive values on a* axis indicate redness while negative values indicate greenness. On the b* axis, the yellow colour is reflected by positive values while blue is represented by negative values. For both axes, zero is neutral grey.

The development of MIRS prediction models was described in detail in **Chapter 2**. Briefly, prediction models were developed separately for each colour index using partial least squares regression analysis (PROC PLS; SAS Institute Inc.) with untreated spectra. Accuracy of prediction was estimated in external validation on 25% of total data, while the remaining 75% was used to calibrate the prediction equations. This procedure was repeated four times using a different validation data set. The external validation correlation coefficient (r_v) was used to define the accuracy of each MIRS predictive model. The highest accuracy of prediction was obtained for b* index ($r_v = 0.72$) while, a* and L* indices were related poorly with the MIR spectrum ($r_v = 0.30$ and $r_v = 0.55$, respectively) (**Chapter 3**).

5.3.2 Data editing

Principal component analysis was undertaken on all 174,062 spectra (PROC PRINCOMP, SAS Institute Inc.). The first four principal components explained 97.33% of the entire spectral variation. Mahalanobis distance was computed; it was defined as the sum of squares of the centered and scaled scores of the p principal components (Brereton, 2015). Mahalanobis distance distribution has a χ^2 distribution shape (with degree of freedom computed on principal components) and a threshold of 97.5% was set up on the curve tails. All spectra out of this area were considered outliers and were deleted. A total of 16,870 records were discarded from all data obtained. Milk colour traits were then predicted by applying the MIRS models developed in **Chapter 2** to the retained spectra. Only DIM between 5 and 305 from parities 1 to 10 were retained. Obvious data set errors (milk yield and milk fat and urea content lower than 2 kg, 2%, and 2% respectively) were deleted. Values for each trait that were >3 standard deviations from the mean were considered outliers and removed. All three milk colour traits were normally distributed. Contemporary group was defined as experimental treatment by test-day on the research farms and herd test-day on the commercial farm. Only contemporary groups with >10 observations were retained. Following all edits, the final data set consisted of 136,807 milk spectra from 16,543 lactations from 9,824 cows.

5.3.3 Data analysis

Spearman rank correlations among the gold standard and predicted milk colour indices (L*, a*, b*), milk yield (kg), milk fat (%), milk protein (%), milk lactose (%), urea (mg/dL), casein (%) and SCC (cells/mL) were computed.

For the purpose of quantifying the effect of stage of lactation or parity on the correlation between traits, stage of lactation was stratified into classes (≤ 60 DIM, from 61 to 159 DIM, and ≥ 160 DIM) and parity was defined as 1, 2, and ≥ 3 parities. Spearman rank correlations among traits were computed within each class and the significance of the differences in the correlations between pairwise classes was determined using the Fisher r-to-z transformation.

Factors associated with each of the three predicted milk colour trait were determined using the following linear mixed animal model in ASREML (Gilmour et al., 2011):

$$Y_{jklmnopqr} = \sum_{i=1}^{j} Breed_j + Het_k + Rec_l + Par_m + DIM_n + Session_o + Month_p + Par_m * DIM_n + PEwithin_q + PEacross_q + Cont\_group_q + e_{jklmnopqr}$$

Where: $Y_{jklmnopqr}$ is the milk colour trait (L*, a*, and b*) MIRS-predicted; $Breed_j$ represents the *j*-th proportion of genes of Friesian, Jersey, Montbeliarde, Norwegian Red and other breeds (proportion of Holstein was not included in the model to avoid linear dependencies) treated as a continuous fixed effect; Het_k is the fixed effect of the *k*-th class of individual heterosis coefficient (12 classes: 0%, 1-10%, 11-20%, 21-30%, 31-40%, 41-50%, 51-60%, 61-70%, 71-80%, 81-90%, 91-99%, 100%); Rec_l is the fixed effect of *l*-th class of the individual recombination loss coefficient (12 classes: 0%, 1-10%, 11-20%, 21-30%, 31-40%, 41-50%, 51-60%, 61-70%, 71-80%, 81-90%, 91-99%, 100%); Par_m is the fixed effect of *m*-th class of parity (5 classes: 1, 2, 3, 4, ≥ 5); DIM_n is the fixed effect of the *n*-th class of stage of lactation (10 classes: 5-30, 31-60, 61-90, 91-120, 121-150, 151-180, 181-210, 201-240, 241-270, 271-305 DIM); $session_o$ is the fixed effect of the *o*-th class of milking time (3 classes: AM, PM, or combined); $month_p$ is the fixed effect of the *p*-th class of month of test (12 classes: January, February, March, April, May, June, July, August, September, October, November, December); Par_m*DIM_n is the fixed effect of the two-way interaction between the *m*-th class of parity and the *n*-th class of stage of lactation; $PEwithin_q$ is the random effect of the within lactation permanent environmental effect of the *q*-th cow where PEwithin ~N(0,I$\sigma^2_{PEwithin}$); $PEacross_q$ is the random effect of the across lactation permanent environmental effect of the *q*-th cow where PEacross ~N(0,I$\sigma^2_{PEacross}$); $Cont\_group_q$ is the random effect of the contemporary group of the *q*-th cow where Cont_group ~N(0,I$\sigma^2_{Cont_group}$); $e_{jklmnopqr}$ is the random effect of the residual where e~N(0,Iσ^2_e). A series of supplementary analysis were undertaken, in which test-day milk yield or milk fat concentration was included in the model as a covariate. Least squares means were derived for a reference animal which was represented by a third parity cow, 100% Holstein, milked in the morning, averaged across all stages of lactation and months of test.

Variance components were estimated for the MIRS-predicted milk colour traits ($\hat{L}$*, â*, and $\hat{b}$*), as well as for milk yield, log_{10}SCC, and concentrations of protein, fat, lactose, urea, and casein using random regression animal models fitted across lactation in ASREML (Gilmour et al, 2011); variance components were restricted to only the 8,519 cows that were ≥75% Holstein-Friesian. The number of test-day records remaining was 98,253 from 14,204 lactations. The data was divided into 10 residual groups based on

DIM as 5 to 30 DIM, 31 to 60 DIM, …, 241-270 DIM, and 271 to 305 DIM. The estimated residual variance, within group, was assumed to be homogenous, but between group the estimated residual variance could be heterogeneous. No residual (co)variance was assumed among residual groups. The model fitted was the same as previously described, with the exception that the effect of month of test was excluded from the analysis, as the effect of contemporary group was fitted as a fixed term. The effect of DIM class was also excluded from the statistical analysis as Legendre polynomials on each individual DIM were fitted as fixed term. Moreover, the animal additive genetic effect was added as a random term where the additive effect followed the assumptions of ~N(0,Aσ_a^2). Legendre polynomials were fitted as a random term on both the additive genetic effect as well as on the within lactation permanent environmental effect. The most parsimonious order of fixed Legendre polynomials was based on visual inspection of the resulting lactation profile for each milk colour trait for the different polynomial orders. In all instances, a cubic Legendre polynomial was the most appropriated as minimal differences were detected between lactation profiles generated with higher order polynomials. Based on the Akaike information criterion, the most parsimonious random covariance function was a cubic polynomial fitted to both the additive genetic and the within-lactation permanent environmental effects for $\hat{L}$* and $\hat{b}$*, while for â* the polynomial order for both random terms was quadratic.

Univariate analyses using ASREML (Gilmour et al., 2011) was carried out also using repeatability animal model on both gold standard and MIRS-predicted milk colour traits, milk yield, milk composition, and $\log_{10}$SCC. The model was the same as described for the phenotypic analyses.

Genetic covariance function coefficients were estimated as $\delta^2 = \Phi K \Phi'$, where δ^2 is the 301 x 301 (co)variance matrix for the MIRS-predicted milk colour trait, milk yield, milk composition and $\log_{10}$SCC, Φ is the 301 x n matrix of Legendre polynomial regressed on DIM, and K is the n x n (co)variance matrix of the additive genetic (or within lactation permanent environment) effect. Standard errors of the heritability estimates were calculated using a Taylor series expansion following Fisher et al. (2004). Pairwise genetic correlations between traits were calculated using a series of bivariate random regression models, fitting the same model as used for the univariate analyses. Residual groups were as defined in the univariate analyses, but within-group residual covariances were estimated. Covariance functions for the random terms were reduced to a

quadratic polynomial in order to meet log-likelihood convergence. Standard error of the genetic correlation were calculated as in Falconer and MacKay (1996).

5.4 Results

Summary statistics for all colour and performance traits are in **Table 5.1**. The mean of the predicted and the respective gold standard milk colour variable was similar. The phenotypic correlations between the milk colour parameters (both gold standard and predicted) and performance traits are in **Table 5.2**.

Table 5.1 Number of records (n), mean, genetic standard deviation (σ_g), heritability (standard error) and repeatability (standard error) for the three gold standard (L*=lightness, a*=redness/greenness, b*=yellowness/blueness) and predicted ($\hat{L}$*, â* and $\hat{b}$*) colour indices as well as milk yield, and concentrations of protein, fat, lactose, urea, and casein, and somatic cell score (SCS = $\log_{10}$ SCC).

Trait	n	Mean	σ_g	Heritability	Repeatability
L*	590	81.60	0.54	0.16 (0.14)	0.29 (0.16)
a*	569	-3.88	0.17	0.07 (0.16)	0.61 (0.13)
b*	594	8.04	0.69	0.13 (0.12)	0.13 (0.12)
$\hat{L}$*	133,611	81.63	0.30	0.29 (0.02)	0.40 (0.01)
â*	133,653	-4.05	0.07	0.10 (0.01)	0.18 (0.00) [a]
$\hat{b}$*	133,528	8.23	0.12	0.35 (0.01)	0.38 (0.01)
Milk yield (kg)	134,155	13.44	1.48	0.21 (0.02)	0.67 (0.00) [a]
Protein (%)	128,561	3.71	0.17	0.46 (0.02)	0.59 (0.01)
Fat (%)	128,256	4.61	0.36	0.29 (0.01)	0.31 (0.01)
Lactose (%)	128,510	4.76	0.09	0.36 (0.02)	0.49 (0.01)
Urea (mg/dL)	127,982	30.59	1.95	0.14 (0.01)	0.25 (0.01)
SCS ($\log_{10}$cells/mL)	75,950	1.79	0.09	0.05 (0.01)	0.44 (0.01)
Casein (%)	128,615	2.81	0.14	0.46 (0.02)	0.59 (0.01)

[a] SE values were not different from zero

A higher L* (i.e. lighter milk) was associated with a more positive a* (i.e. more red) and more positive b* (i.e. more yellow) values. Both a* and b* were not correlated with each other when derived from the MIRS prediction equations despite a weak negative correlation (-0.11) between the gold standard a* and b* values. Milk yield was negatively correlated with all colour traits predicted from MIRS but was only negatively correlated with L* and b* values when the gold standard values were used (-0.42 and -0.54, respectively). Milk fat concentration was moderately correlated with both gold standard (0.43) and predicted (0.58) L* as well as being positively correlated with gold

standard and predicted b* (0.64 to 0.77). Milk protein concentration was moderately positively correlated with both L* (0.34 to 0.50) and b* (0.50 to 0.62). Similarly, casein concentration in the milk was moderately positively correlated with both L* (0.35 to 0.52) and b* (0.46 to 0.57).

Several of the phenotypic correlations between at least one of the colour traits and the performance traits, however, differed by parity (**Table 5.3**) or stage of lactation (**Table 5.4**); the exceptions were the correlations between all colour traits with milk fat concentration, milk protein concentration, and milk casein concentration which did not differ by stage of lactation (**Table 5.4**).

Jersey breed had the highest fat concentration (5.10%, SE=0.04) compared to Holstein (3.90%, SE=0.03), Friesian (3.99%, SE=0.06), Norwegian Red (3.94%, SE=0.07) and Montbeliarde (3.62%, SE=0.09).

5.4.1 Lightness colour (L*)

Milking-time (P<0.001), stage of lactation (P<0.001), Jersey proportion (P<0.001), month of the year (P<0.001), parity (P<0.001), the interaction between parity and stage of lactation (P<0.001), Montbeliarde proportion (P<0.001), recombination loss (P<0.001), Friesian proportion (P<0.05), and heterosis (P<0.05) were all associated with $\hat{L}$*; $\hat{L}$* was not associated with the proportion of Norwegian Red in the animals. $\hat{L}$* generally increased as the calendar year progressed although a dip in $\hat{L}$* was evident in August (**Figure 5.1**).

Mean $\hat{L}$* was 81.25 (SE= 0.02) in the morning milking and 81.89 (SE=0.02) in the evening milking. The milk of Jerseys had higher $\hat{L}$* values than Holsteins, Friesians (P<0.001), Norwegian Reds (P<0.001) and Montbeliarde cows (P<0.001) (**Table 5.5**). The regression coefficient of $\hat{L}$* on Jersey breed proportion changed from 0.41 to -0.12 following the adjustment for difference in milk fat concentration in the model (**Table 5.5**). Adjustment for differences in milk yield in the model had a minimal effect on the regression coefficient of $\hat{L}$* on Jersey proportion. Although the trend of $\hat{L}$* across lactation differed statistically (P<0.001) by parity, the biological impact of the interaction was minimal (**Figure 5.2**). Irrespective of parity, $\hat{L}$* consistently increased with advancing stages of lactation. Mean $\hat{L}$* for parity 1 was 81.36 (SE=0.02), for parity 2 was 81.32 (SE=0.03), for parity 3 was 81.29 (SE=0.03), for parity 4 was 81.26 (SE=0.03), for parity 5+ was 81.22 (SE=0.03) (**Figure 5.2**).

The heritability of gold standard L* estimated with the repeatability model was 0.16 (SE=0.14), while the repeatability was 0.29 (SE=0.16; **Table 5.1**). Heritability estimates for $\hat{L}$* calculated using random regression models ranged between 0.15 ± 0.02 (30 DIM) and 0.46 ± 0.02 (210 DIM; **Figure 5.7**). Within trait genetic correlations approached unity between adjacent DIM; all within trait genetic correlations were positive, and had a minimum of 0.02 ± 0.02 between 5 and 305 DIM (**Figure 5.8).** On average, $\hat{L}$* was negatively genetically correlated with milk yield (-0.65 ± 0.02 to -0.37 ± 0.06 at 249 and 37 DIM, respectively), milk lactose concentration (-0.34 ± 0.05 to 0.07 ± 0.06 at 305 and 41 DIM, respectively) and milk urea content (-0.18 ± 0.03 to -0.10 ± 0.11 at 252 and 5 DIM, respectively; **Figure 5.9**). Positive genetic correlations existed between $\hat{L}$* and both milk fat concentration (0.32 ± 0.09 to 0.78 ± 0.01 at 5 and 249 DIM, respectively) and milk protein concentration (0.43 ± 0.07 to 0.91 ± 0.01 at 5 and 305 DIM, respectively; **Figure 5.9**).

5.4.2 Red-green colour (a*)

Factors associated with â* included milking-time (P<0.001), month of the year (P<0.001), stage of lactation (P<0.001), parity (P<0.001), Jersey proportion (P<0.001), the two-way interaction parity-by-stage of lactation (P<0.001), Norwegian Red proportion (P<0.001), heterosis (P<0.001), recombination loss (P<0.001), and Friesian proportion (P<0.02); the proportion of Montbeliarde in the cow was not associated with â* values. The â* colour of milk was relatively consistent across months of the year with a peak (i.e., more red) in August (-3.55) and a minimum (i.e., more green) of between -4.46 to -4.43 between March and June (**Figure 5.1**). Mean â* was -3.95 (SE= 0.01) in the morning milking and -3.82 (SE=0.01) in the evening milking. The milk of Friesians, Jerseys, Norwegian Reds and Montbeliardes was more green (i.e., lower â*) than that of Holsteins (**Table 5.5**).

5.4.3 Yellow-blue colour (b*)

Milking time (P<0.001), Jersey proportion (P<0.001), parity (P<0.001), stage of lactation (P<0.001), month of the year (P<0.001), the two-way interaction parity-by-stage of lactation (P<0.001), recombination loss (P<0.001), and Montbeliarde proportion (P<0.01), were all associated with $\hat{b}$*; $\hat{b}$* was not associated with either the proportion of Friesian and Norwegian Red nor the heterosis coefficient of the cow. There was a general trend for the $\hat{b}$* value of milk to increase with calendar month (**Figure 5.1**) varying from

6.69 (in January, SE=0.18) to 8.47 (in December, SE=0.15). Including fat concentration as a covariate in the statistical model did not greatly alter the trend across months (**Figure 5.1**). The mean unadjusted $\hat{b}^*$ value was 7.72 (SE=0.04) and 8.89 (SE=0.04) in morning and evening milking, respectively. Jersey cows had more yellow milk than Holsteins with a $\hat{b}^*$ value of + 1.85 (+ 0.49 after adjustment for milk fat content) relative to a $\hat{b}^*$ value of 0 for Holstein cows. Otherwise, the milk of Montbeliarde cows was, on average, bluer than the milk of Holsteins (**Table 5.5**). Milk $\hat{b}^*$ value was influenced by recombination loss only in animals with a gene recombination percentage between 30% and 80%.

Although a significant interaction between parity and stage of lactation existed for the association with $\hat{b}^*$, the trend in $\hat{b}^*$ across lactation was nonetheless similar across parities decreasing from between 5 and 30 days in milk to between 31 and 60 days in milk and increasing thereafter (**Figure 5.4**). Mean $\hat{b}^*$ in parity 1 animal was lowest (7.36) while mean $\hat{b}^*$ in second parity animals was 7.56; the mean $\hat{b}^*$ of older parity animals were similar (7.80 to 7.87).

Including milk yield in the statistical model did not impact the lactation profile for $\hat{b}^*$ but the difference between parities increased (**Figure 5.5**); for example the mean difference in $\hat{b}^*$ between parity 1 and parity 3 animals increased from 0.43 without milk yield in the model to 0.60 with milk yield in the model. Including milk fat concentration in the statistical model altered the shape of the lactation profiles for $\hat{b}^*$ with no observed reduction in $\hat{b}^*$ in early lactation but also a widening of the difference in $\hat{b}^*$ between parity one and older parity animals especially in early lactation (**Figure 5.6**).

The heritability and repeatability estimates calculated using the repeatability animal model for gold standard b* was 0.12 (SE=0.13) and 0.13 (SE=0.12), respectively (**Table 5.1**). The heritability and repeatability estimates for the $\hat{b}^*$ parameter estimated using a repeatability model that phenotypically adjusted for milk fat concentration was 0.25 (SE=0.01) and 0.32 (SE=0.01), respectively; the genetic standard deviation of $\hat{b}^*$ following the genetic adjustment for milk fat concentration was 0.25 (coefficient of genetic variation of 3.22%). Heritability estimates from the random regression analysis of $\hat{b}^*$ varied between 0.18 ± 0.02 (21 DIM) to 0.56 ± 0.03 (305 DIM; **Figure 5.7**). Within trait genetic correlations weakened as the time between DIM increased, and had a minimum of 0.32 ± 0.02 between 5 and 305 DIM (**Figure 5.8**). Milk $\hat{b}^*$ was genetically positively correlated with $\hat{L}^*$ (0.31 ± 0.08 to 0.74 ± 0.02 at 5 and 293 DIM, respectively), milk fat concentration (0.82 ± 0.03 to 0.96 ± 0.01 at 5 and 305 DIM, respectively), and

milk protein concentration (0.58 ± 0.06 to 0.83 ± 0.01 at 8 and 305 DIM, respectively; **Figure 5.11).** Negative genetic correlations existed between $\hat{b}^*$ and â* (-0.47 ± 0.07 to 0.01 ± 0.07 at 10 and 297 DIM, respectively), milk yield (-0.62 ± 0.02 to -0.45 ± 0.05 at 220 and 14 DIM, respectively), and milk lactose concentration (-0.43 ± 0.04 to -0.08 ± 0.04 at 305 and 62 DIM, respectively; **Figure 5.11**).

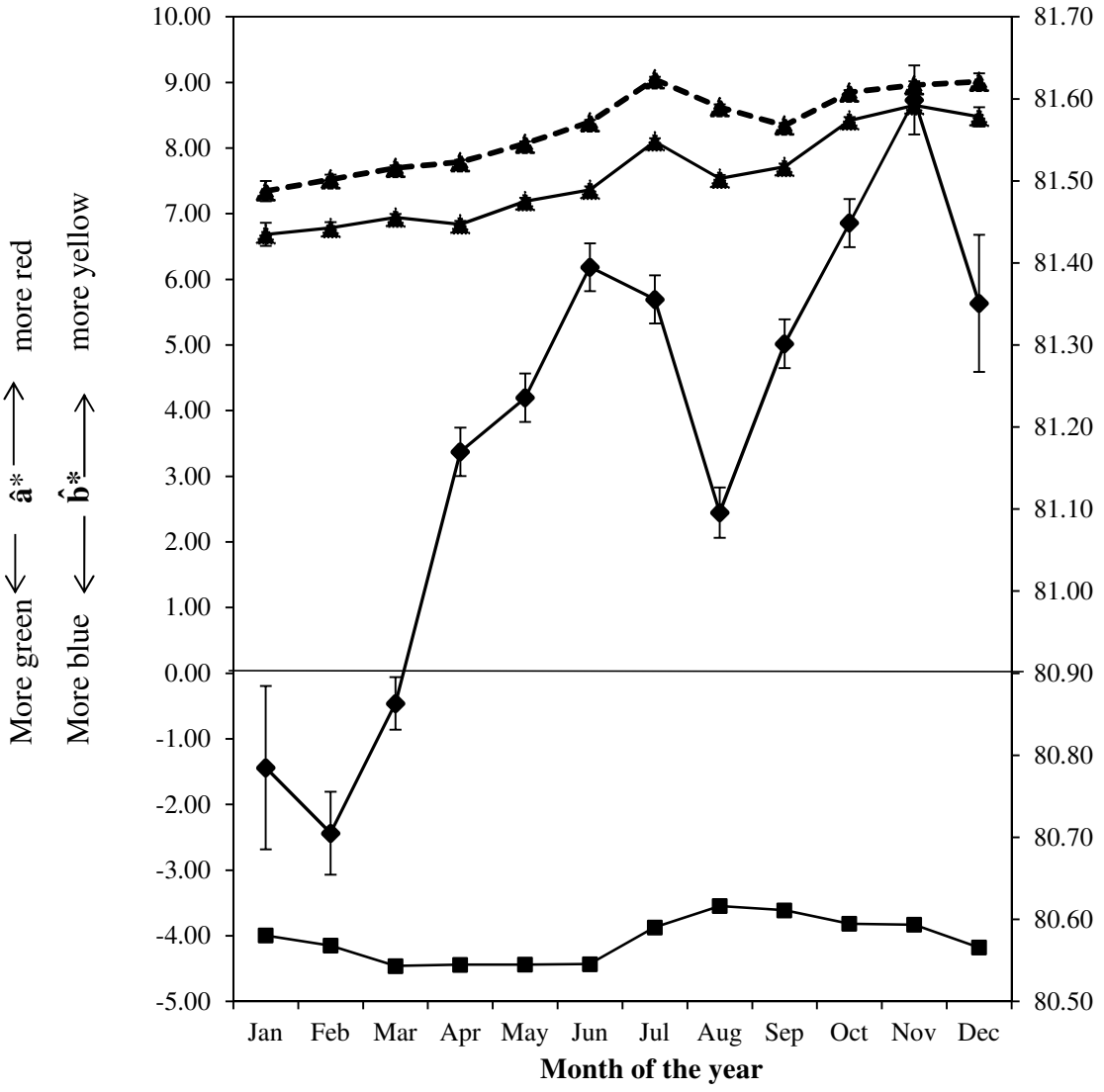


Figure 5.1 Monthly least-squares means along the calendar year of $\hat{L}^*$ (—♦—; on the secondary vertical axis), â* (—■—; on the primary vertical axis), $\hat{b}^*$ (—▲—; on the primary vertical axis); $\hat{b}^*$ considered fat concentration (4.61%) as a covariate in the model (---▲---; on the primary vertical axis) (average SE = 0.05).

Table 5.2 Spearman rank correlations[1] between the colour traits and milk yield, concentrations of protein, fat, lactose, urea, and casein, and SCC. Correlations with gold standard values of L* (n=590), a* (n=562), and b* (n=597) are above the diagonal and correlations with predicted values are below the diagonal.

Item	L*	a*	b*	Milk yield	Protein	Fat	Lactose	Urea	SCC	Casein
L*	-	0.32	0.55	-0.42	0.34	0.43	-0.17	0.06	0.00	0.35
a*	0.24	-	-0.11	0.05	-0.22	-0.05	-0.01	-0.18	0.15	-0.23
b*	0.74	0.00	-	-0.54	0.50	0.64	-0.33	0.32	0.05	0.46
Milk yield	-0.64	-0.18	-0.69	-	-0.42	-0.54	0.29	-0.31	-0.04	-0.45
Protein	0.50	-0.20	0.62	-0.44	-	0.48	-0.38	0.43	0.21	0.93
Fat	0.58	0.04	0.77	-0.57	0.51	-	-0.18	0.19	0.02	0.51
Lactose	-0.18	0.01	-0.45	0.33	-0.31	-0.22	-	-0.42	-0.06	-0.24
Urea	0.05	-0.21	0.42	-0.32	0.37	0.18	-0.42	-	0.26	0.39
SCC	0.02	0.05	-0.01	-0.03	0.03	-0.11	0.08	0.07	-	0.20
Casein	0.52	-0.14	0.57	-0.46	0.92	0.51	-0.16	0.31	0.02	-

[1] Correlations < |0.07| were not different from zero (P > 0.05).

Table 5.3 Spearman rank correlations between the three gold standard milk colour indices (L*, a* and b*) and milk yield, concentrations of protein, fat, lactose, urea, and casein, and SCC in different parities.

	L*			a*			b*		
Parity	1	2	≥ 3	1	2	≥ 3	1	2	≥ 3
Milk yield (kg)	-0.39	-0.43	-0.40	0.09[a]	-0.13[b]	0.06[a]	-0.43[a]	-0.48[a]	-0.70[b]
Protein (%)	0.36	0.28	0.34	-0.18	-0.17	-0.21	0.44[a]	0.42[a]	0.58[b]
Fat (%)	0.42	0.35	0.51	0.07	-0.01	-0.10	0.54[a]	0.55[a]	0.74[b]
Lactose (%)	-0.16[a]	-0.02[a]	-0.33[b]	0.06	0.06	-0.08	-0.32	-0.29	-0.38
Urea (mg/dL)	0.09[a]	-0.10[b]	0.17[a]	-0.21	-0.13	-0.09	0.30[a]	0.21[b]	0.40[a]
SCC (cell/mL)	0.08	0.07	-0.19	0.11	0.27	0.08	0.09	0.19	-0.07
Casein (%)	0.36	0.31	0.36	-0.20	-0.16	-0.20	0.38[a]	0.38[a]	0.55[b]

[a-b] Correlations within the same row with different superscripts are different ($P < 0.05$) from each other.

Table 5.4 Spearman rank correlation between the three milk colour indices (L*, a* and b*) and milk yield, concentrations of protein, fat, lactose, urea, and casein, and SCC in different stages of lactation.

	L*			a*			b*		
Days in milk	0-60	61-159	160-305	0-60	61-159	160-305	0-60	61-159	160-305
Milk yield	0.04[b]	-0.50[a]	-0.35[a]	-0.07[a]	0.11[b]	-0.12[a]	-0.29[a]	-0.59[b]	-0.41[a]
Protein	0.37	0.37	0.27	-0.21	-0.20	-0.17	0.35	0.40	0.43
Fat	0.45	0.54	0.37	0.13	-0.002	-0.01	0.67	0.69	0.54
Lactose	-0.14	-0.04	-0.16	-0.24[a]	-0.07[a]	0.06[b]	-0.02[a]	-0.20[a]	-0.28[b]
Urea	-0.20[a]	0.13[b]	-0.03[a]	-0.07	-0.12	-0.12	-0.14[a]	0.31[b]	0.06[a]
SCC	0.10	0.05	-0.18	0.13	0.15	0.27	-0.05[a]	0.17[a]	-0.33[b]
Casein	0.36	0.36	0.27	-0.28	-0.21	-0.12	0.35	0.33	0.41

[a-b] Correlations within the same row with different superscripts are different ($p<0.05$) from each other.

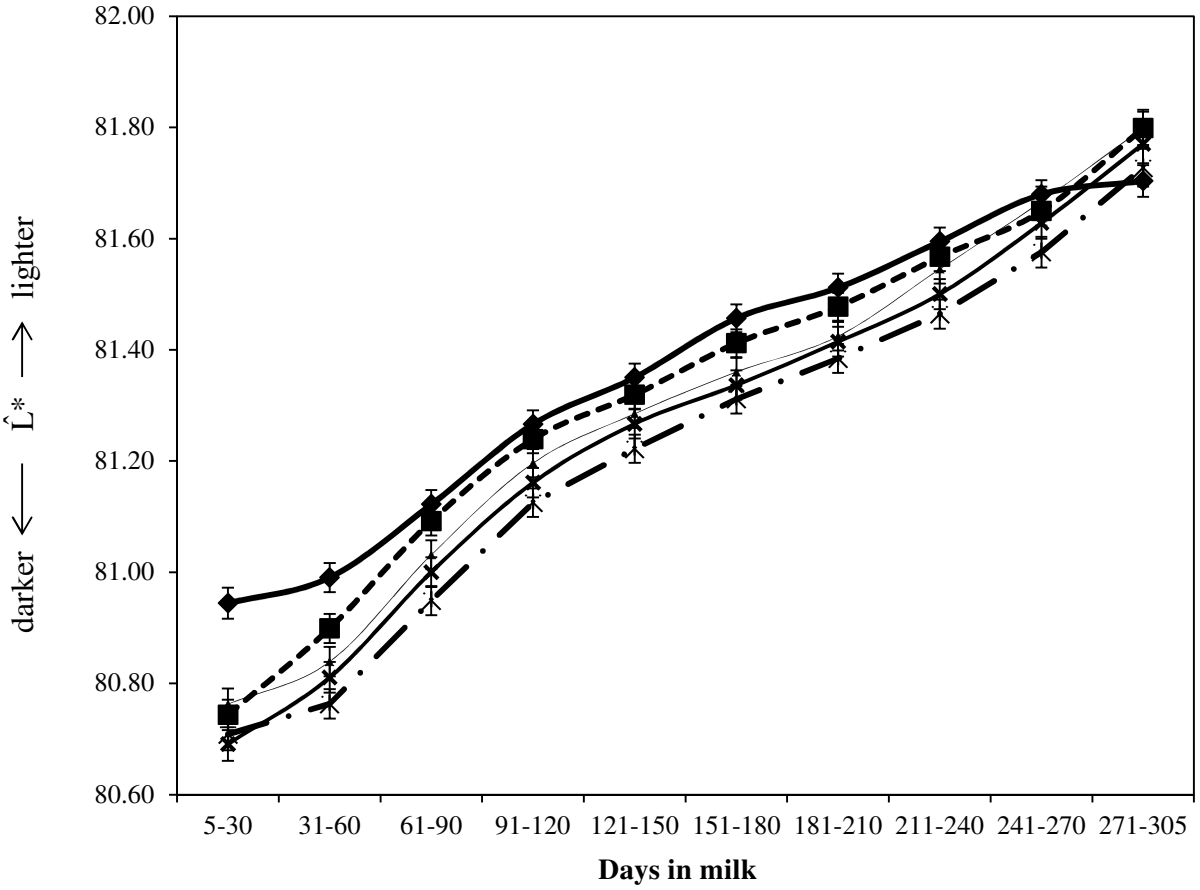


Figure 5.2 Least squares means of L̂* values throughout lactation in parity 1 (—♦—), parity 2 (---■---), parity3 (—▲—), parity 4 (—X—), and parity ≥ 5 (-··X-··-) (average SE = 0.03).

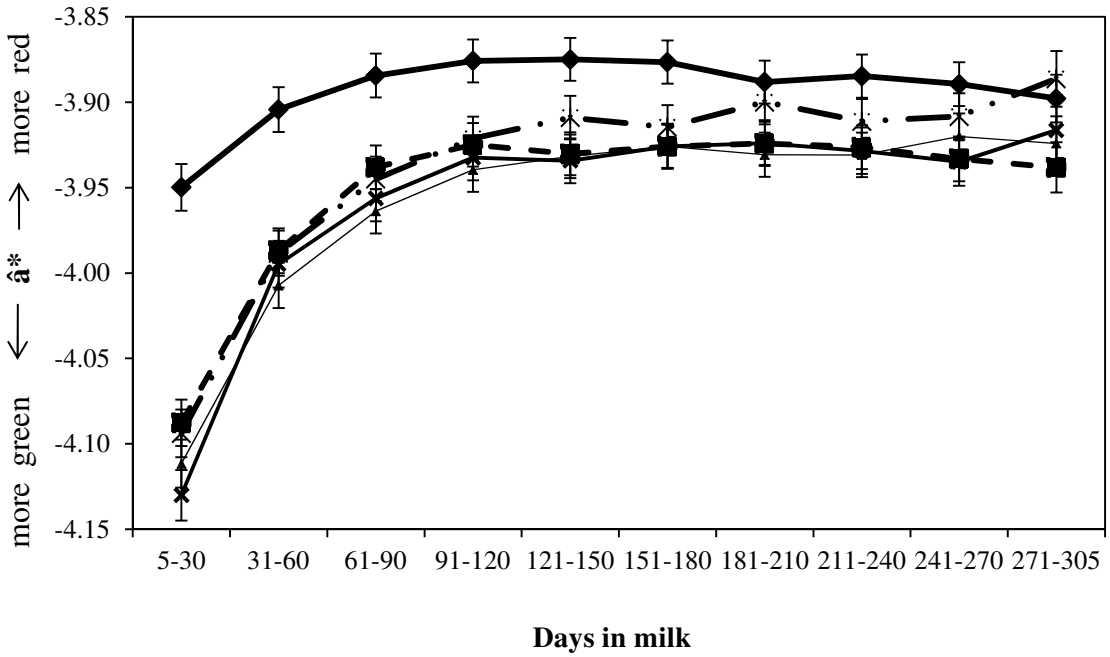


Figure 5.3 Least squares means of â* values throughout lactation in parity 1 (—♦—), parity 2 (---■---), parity3 (—▲—), parity 4 (—X—), and parity ≥ 5 (-··X··-) (average SE = 0.01).

The shape of the lactation profile differed by cow parity number although, within parity, the lowest â* value was in early lactation reaching a plateau from mid-lactation on (**Figure 5.3**). The profile of first lactation cows differed biologically from that of later parity cows which in turn were similar to each other. Across lactation the mean â* of first parity cows was -3.89 compared to a parity mean of between -3.94 and -3.96 for later parity cows (**Figure 5.3**).

The heritability and repeatability estimates of gold standard a* (calculated by the repeatability animal model) was 0.07 (SE=0.18) and 0.60 (SE=0.13), respectively (**Table 5.1**). Heritability values for â* estimated using the random regression models ranged from 0.09 ± 0.01 (30 DIM) to 0.15 ± 0.02 (305 DIM) increasing almost consistently as lactation progressed (**Figure 5.7**). Within trait genetic correlations had a minimum of 0.44 ± 0.02 occurring between 5 and 219 DIM (**Figure 5.8).** Milk â* values were positively genetically correlated with both L̂* (0.24 ± 0.05 to 0.46 ± 0.08 at 97 and 5 DIM, respectively) and $\log_{10}$SCC (0.14 ± 0.09 to 0.46 ± 0.16 at 94 and 5 DIM, respectively), but were negatively genetically correlated with all other milk quality traits (**Figure 5.10**).

Table 5.5 Linear regression coefficients of L̂*, â* and b̂* on breed fractions, and with fat concentration (4.61%) or milk yield (13.44 kg) also included as a covariate in the statistical model.

	L̂*	â*	b̂*
Friesian	-0.13 (0.05)	-0.04 (0.02)	0.00 (0.09)
Jersey	0.41 (0.04)	-0.14 (0.01)	1.85 (0.07)
Norwegian Red	-0.09 (0.07)	-0.07 (0.02)	0.09 (0.11)
Montbeliarde	-0.30 (0.08)	-0.04 (0.03)	-0.33 (0.13)
Adjustment for fat concentration			
Friesian	-0.17 (0.04)	-0.04 (0.02)	-0.09 (0.05)
Jersey	-0.12 (0.04)	-0.19 (0.02)	0.49 (0.04)
Norwegian Red	-0.15 (0.06)	-0.08 (0.02)	0.02 (0.06)
Montbeliarde	-0.19 (0.06)	-0.02 (0.03)	-0.04 (0.07)
Adjustment for milk yield			
Friesian	-0.17 (0.05)	-0.04 (0.02)	-0.06 (0.09)
Jersey	0.34 (0.04)	-0.14 (0.01)	1.74 (0.07)
Norwegian Red	-0.14 (0.07)	-0.08 (0.02)	0.01 (0.11)
Montbeliarde	-0.31 (0.08)	-0.04 (0.03)	-0.31 (0.13)

Table 5.6 Average genetic correlations calculated by random regression models between the three predicted colour traits (L̂*, â* and b̂*) and milk yield, concentrations of protein, fat, lactose, urea, and casein, and $\log_{10}$(SCC/1,000).

	L̂*	â*	b̂*
â*	0.32		
b̂*	0.66	-0.19	
Milk yield	-0.54	-0.08	-0.56
Fat	0.70	-0.17	0.91
Protein	0.77	-0.12	0.71
Lactose	-0.10	-0.32	-0.21
Urea	-0.14	-0.14	0.10
$\log_{10}$(SCC/1,000)	0.11	0.25	-0.01
Casein	0.76	-0.14	0.70

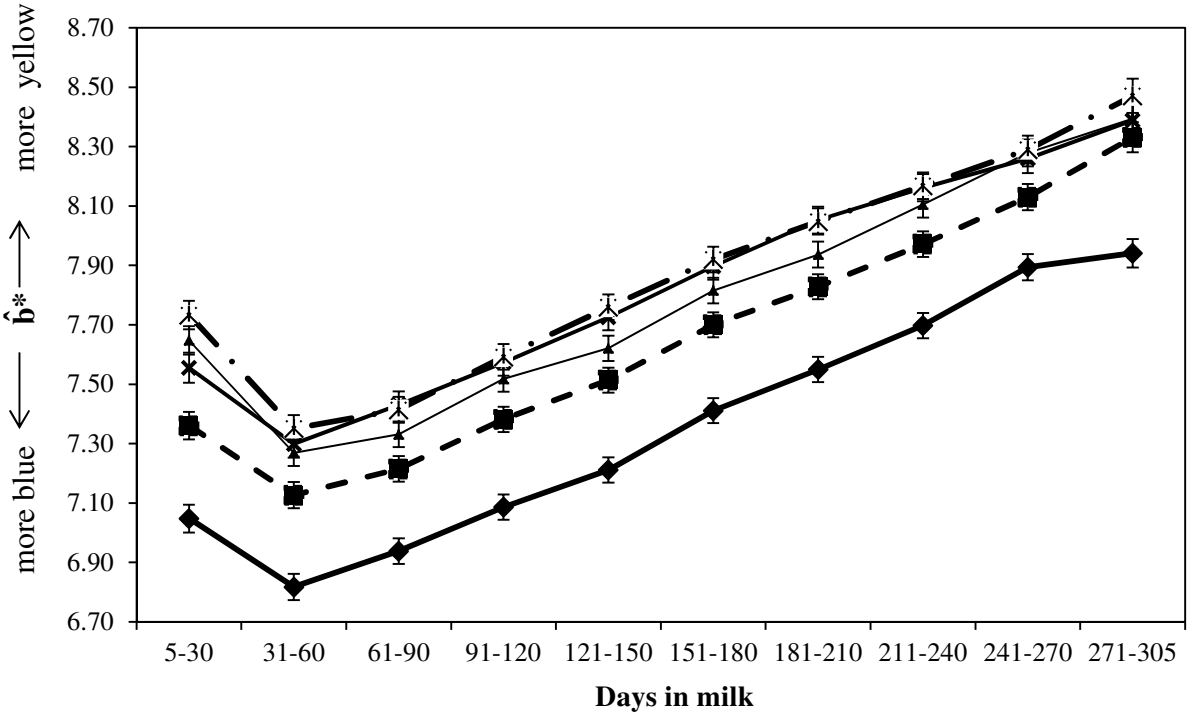


Figure 5.4 Least squares means of $\hat{b}^*$ values throughout lactation in parity 1 (—♦—), parity 2 (---■---), parity3 (—▲—), parity 4 (—X—), and parity ≥ 5 (-··X-··-) (average SE = 0.05).

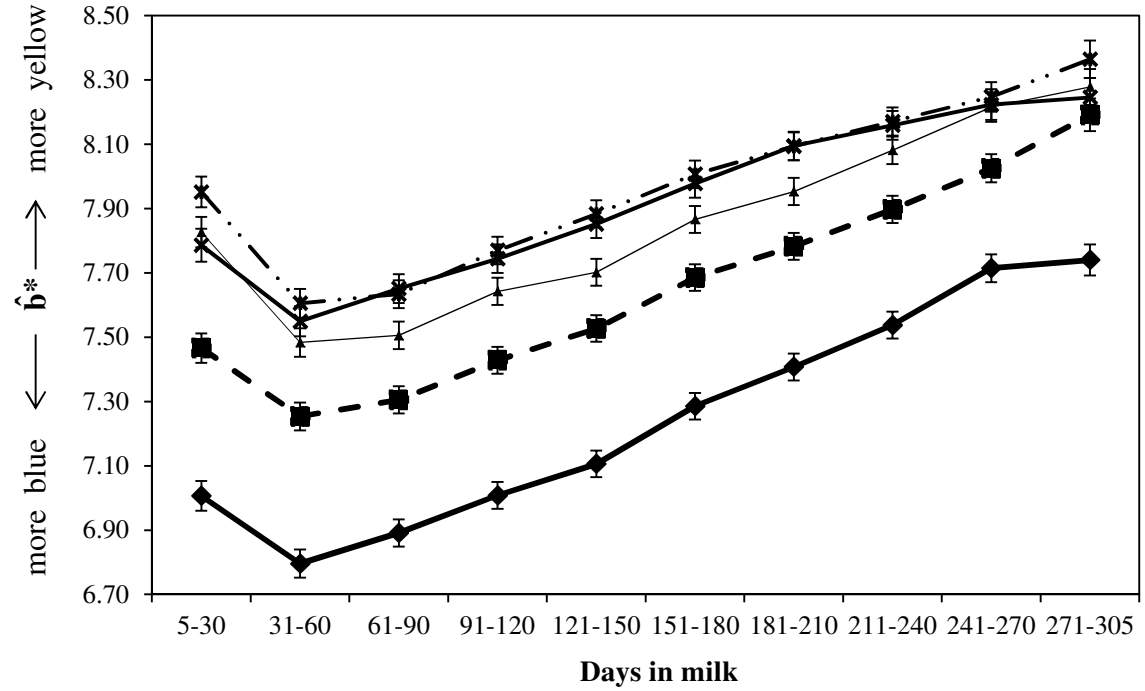


Figure 5.5 Least squares means of $\hat{b}^*$ values throughout lactation in parity 1 (—♦—), parity 2 (---■---), parity3 (—▲—), parity 4 (—X—), and parity ≥ 5 (-··X-··-), with milk yield included as a covariate in the model (fixed milk yield = 13.44kg, average SE = 0.04).

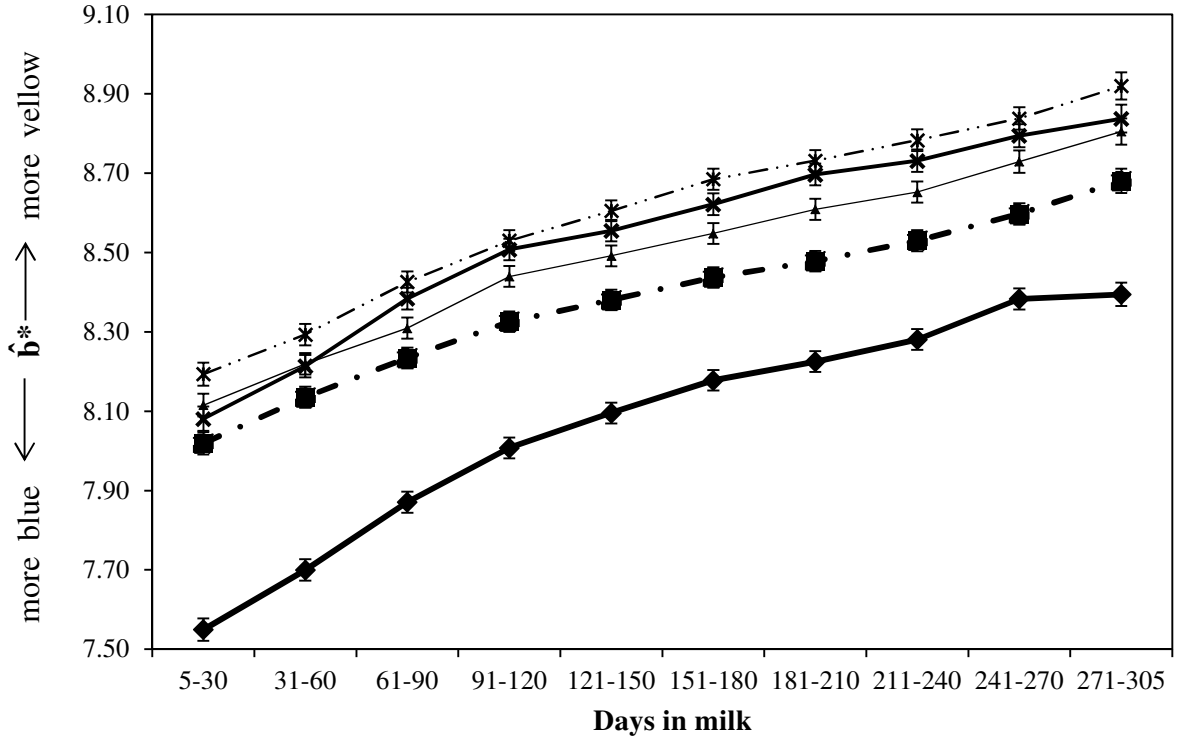


Figure 5.6 Least squares means of $\hat{b}^*$ values throughout lactation in parity 1 (—♦—), parity 2 (---■---), parity3 (—▲—), parity 4 (—X—), and parity ≥ 5 (-··X-··-), with fat concentration included as a covariate in the model (fat concentration fixed = 4.61%, average SE = 0.03).

5.5 Discussion

The objective of the present study was to quantify the contribution of cow-level genetic and non-genetic factors to the observed variability in predicted milk colour as described by lightness (L*), greenness-redness (a*) and blueness-yellowness (b*) indices. The practical implication from this research is to understand, and therefore, predict the possible future changes in milk colour (e.g., with stage of lactation) and therefore facilitate action (e.g., at the processor level) to ameliorate the change in developed products to suit the expected milk colour. For example, different markets demand dairy products (e.g., milk, cheese and butter) differing in colour (Morris et al., 2002; Descalzo et al., 2012). Results from the present study clearly identified genetic and non-genetic factors strongly associated with all three aspects of milk colour. Of particular interest was the existence of considerable genetic variability in each of the three colour parameters; coupled with the ability to predict the parameters from milk MIRS (**Chapter 3**) this suggests that breeding programs to alter milk colour are possible.

Cows included in the present study originated from Irish herds only where the basal diet of the cows was grazed grass, reflective of the production system in Ireland. Mean L*, a*, and b* indices in the present study were slightly different from results reported by Solah et al. (2007) based on Holstein-Friesian cows in Western Australia. These differences were expected as Solah et al. (2007) reported milk fat colour or butter colour instead of milk colour, like in the present study.

5.5.1 Non-genetic factors associated with milk colour

Many studies have heretofore reported associations between parity, stage of lactation and sometimes their interaction on a range of milk production related traits such as milk yield (Sklan et al., 1994), fat and protein concentration (Morris et al, 2002; Jadhav et al., 2008), and somatic cell count (Quist et al., 2008) in dairy cows. Based on the results from the present study, obvious differences among parities and lactation stages also exist for milk colour corroborated by a change in its correlation among traits by stage of lactation and parity.

The time of milking as well as milking frequency have both been documented to affect milk yield (Everett and Wadell, 1970; Gilbert et al., 1973; Erdman and Varner, 1995) and milk composition not only in terms of milk fat and protein concentrations (Quist et al., 2008) but also fatty acids profile (Klei et al., 1997; Ferlay et al., 2010) in dairy cows as well as in dairy ewes (Ploumi et al., 1998).

The present study corroborated the difference in milk colour traits between morning and evening milk. The combination of both greater milk yield (Ouweltjes, 1998) and reduced milk fat concentration (Quist et al., 2008) in morning milk is a reasonable explanation of a less yellow milk colour of morning milk. The adjustment for fat concentration and milk yield had an effect on $\hat{b}^*$ value even in this case, where morning milk had higher values than evening milk supporting the strong correlation between $\hat{b}^*$ and milk fat concentration.

Feeding and herd management, in terms of pasture grazing period, were reported to have a large influence on milk composition especially on milk β-carotene amount (Noziere et al., 2006; gabriel et al., 2012) and milk fatty acid composition (Descalzo et al., 2012). Milk $\hat{L}^*$ and $\hat{b}^*$ increased in colder months (October, November and December) in the present study, which is in agreement with the higher milk colour intensity in cooler seasons reported by Walker et al. (2013). Seasonal variation causes a variation in grazing pasture composition (Hutton et al., 1969; Hall, 1970), and this is

probably the cause of milk composition variation and eventually in milk colour traits. Another factor associated with milk colour traits is udder health as suggested by Espada and Vijverberg (2002) who reported more reddish colour of bovine milk in the presence of mastitis attributable to *Streptococcus Esculin*. Such a result is consistent with the correlation between SCC (and $\log_{10}$SCC) and a* at both phenotypic and genetic level.

5.5.2 Genetics of milk colour

To our knowledge, no studies have documented heterosis and recombination loss effects on milk colour. In the present research, neither heterosis nor recombination loss among breeds had any significant effect on milk colour in spite of observed breed effects. Heterosis effect was also analysed for colour traits considering both the adjustment for fat concentration and milk yield, but heterosis values were still not significant. A possible hypothesis could be that genes that code for milk colour are regressive, but other studies are required on this aspect. The observed significant breed effect on milk colour corroborates previous studies (Winkelman et al., 1999; Berry et al., 2009); the milk of Jersey cows had the highest $\hat{b}$* values, even after adjusting for fat concentration. This could be physiologically explained by both the ability of the cow to convert carotene into vitamin A (Jadhav et al., 2008) as well as the higher fat concentration present in Jersey milk relative to Friesian (Auldist et al., 2004), Holstein (Morales et al., 2000), Montbeliarde (Soyeurt et al., 2006) and Norwegian Red (as previously reported) cows.

The within-trait genetic correlations between each pairwise DIM were all positive with the weakest genetic correlations being generally only between DIM at both peripheries of the lactation. Moreover, the genetic correlations between all other performance traits at the same DIM were relatively consistent across all DIM. With the exception of the â* parameter, however, the heritability estimates for milk colour did change throughout lactation with a tendency to reflect the trend in genetic variance over DIM. Nonetheless, all parameter estimates suggest minimal loss of information if genetic evaluations were undertaken using a repeatability model but also that the ability to dramatically alter the lactation profile of milk colour genetically is low.

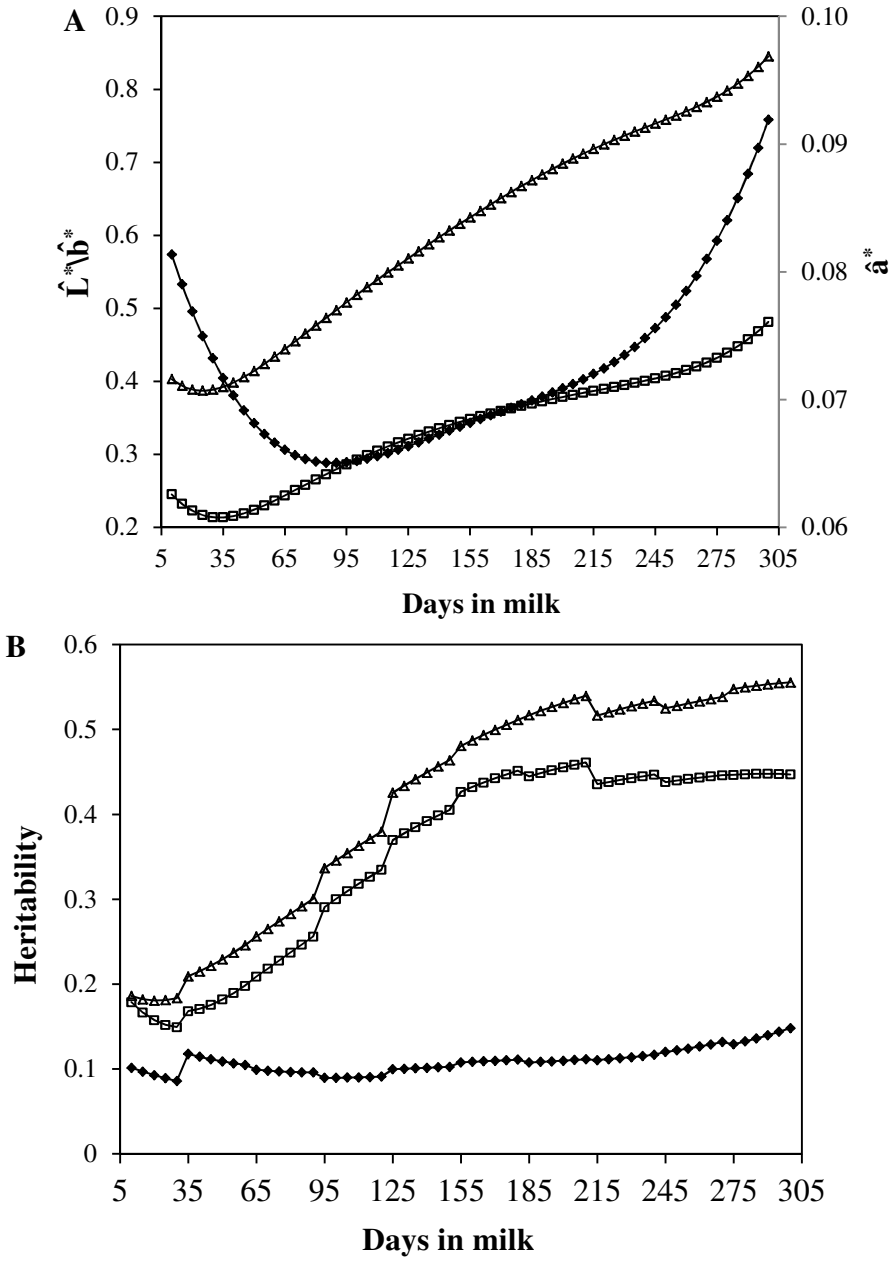


Figure 5.7 A) Genetic standard deviation (SE in parenthesis) for $\hat{L}^*$ (—□—; 0.002 to 0.012), $\hat{a}^*$ (—♦—; 0.02 to 0.07), and $\hat{b}^*$ (—Δ—; 0.01 to 0.03), and B) heritability estimates (SE in parenthesis) for $\hat{L}^*$ (—□—; 0.02 to 0.03), $\hat{a}^*$ (—♦—; 0.01 to 0.02), and $\hat{b}^*$ (—Δ—; 0.01 to 0.03).

The heritability of the three milk colour parameters estimated in the present study agree with previous studies which considered milk colour and fat colour in dairy cows (Winkelman et al., 1999) as well as milk carotenoid concentration in both milk and milk fat of dairy cows (Morris et al., 2002). The coefficient of variation for all three colour traits estimated using the repeatability model was however small (0.37% to 6.88% for the predicted traits) and less than for the other performance traits, such as milk yield (9.81%). The coefficient of variation for all three colour traits was however small (0.4% to 2% for the predicted traits) and less than for the other performance traits. This indicates that relatively few records are required to achieve high accuracy of selection for these traits but the lack of considerable genetic variation suggests that actually achieving genetic gain may prove difficult. This could be exacerbated by the presence of a moderate negative genetic correlation averaged across all DIM between milk yield and both $\hat{L}^*$ (-0.54) and $\hat{b}^*$ (-0.56), manifesting itself as a requirement to place emphasis on both milk colour parameters to avoid any change in milk colour as a repercussion of selection for greater milk yield as exists in most breeding goals (Miglior et al., 2005).

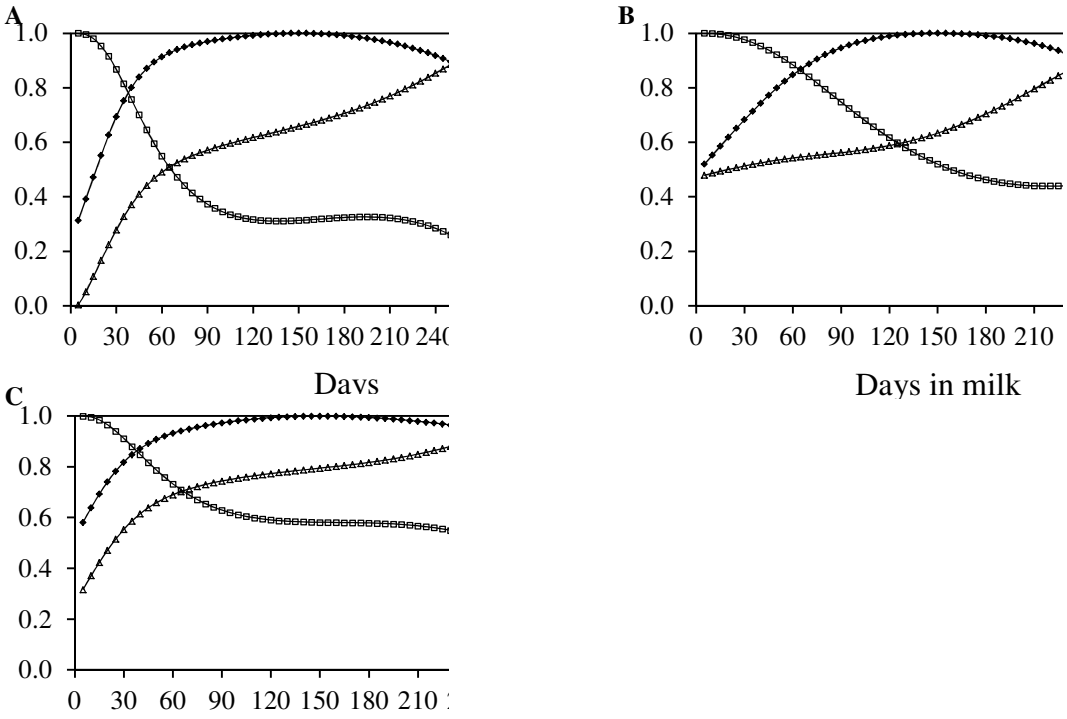


Figure 5.8 Within trait genetic correlations between 5 DIM (—□—), 150 DIM (—♦—), and 305 DIM (—Δ—) and the rest of lactation for □) $\hat{L}^*$; B) â*; C) $\hat{b}^*$. Standard errors ranged between 0.00 and 0.02.

The negative genetic correlation between $\hat{b}^*$ (yellow-blue) index and milk yield agrees with previous studies (Winkelman et al., 1999; Morris et al., 2002) and could be an artefact of dilution of colour with greater milk yield. The heritability and genetic standard deviation for $\hat{b}^*$ estimated using a repeatability animal model that was phenotypically adjusted for milk yield was 0.33 (SE=0.02) and 0.36 (SE=0.01), respectively, indicating a reduction in genetic variability in $\hat{b}^*$ phenotypically independent of milk yield. The coefficient of genetic variation for $\hat{b}^*$ following genetic adjustment for difference in milk yield was 6.53% (i.e., 94% of the origin genetic variation). Using the heritability and repeatability estimates from the repeatability animal model, each genetic standard deviation unit increase in milk yield though breeding for milk yield alone is expected to reduce $\hat{b}^*$ by 0.19. Therefore to hold the $\hat{b}^*$ colour of milk constant following single trait selection on milk yield would require a relative emphasis of 33% on milk $\hat{b}^*$ colour; the gain in milk yield with such an index would be 0.81 times that of the gain in milk yield where only milk yield constituted the breeding goal. Hence, attempts to halt any change in milk colour due to breeding programs for increased milk production may require milk colour to be included in the breeding goal with some emphasis which will have repercussions in genetic gain for milk yield and other traits in the breeding goal.

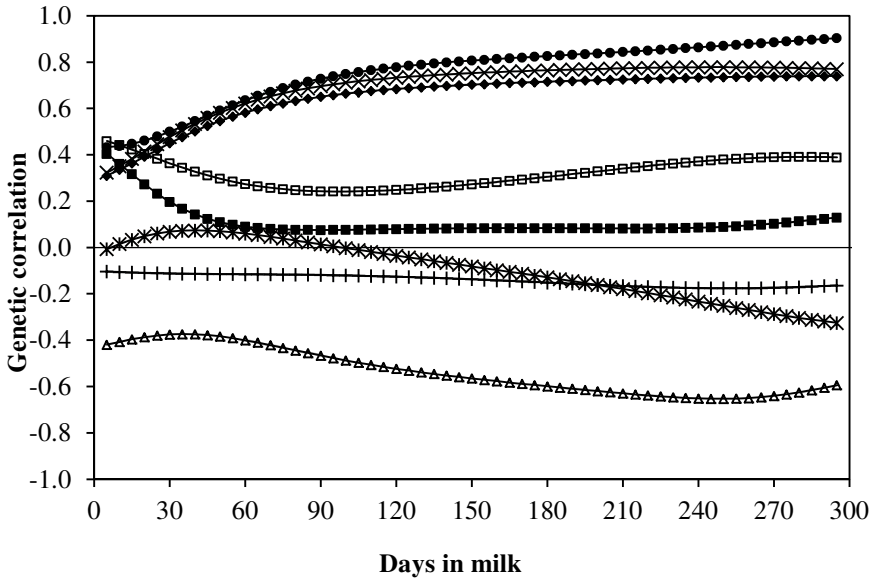


Figure 5.9 Genetic correlations (SE in parenthesis) between $\hat{L}^*$ and $\hat{a}^*$ (—□—; 0.04 to 0.08), $\hat{b}^*$ (—♦—; 0.01 to 0.08), milk yield (—Δ—; 0.02 to 0.06), protein concentration (––●—; 0.01 to 0.07), fat concentration (—×—; 0.01 to 0.09), lactose concentration (—*—; 0.02 to 0.10), urea concentration (—+—; 0.03 to 0.11), and $\log_{10}$(SCC/1,000) (—■—; 0.05 to 0.17).

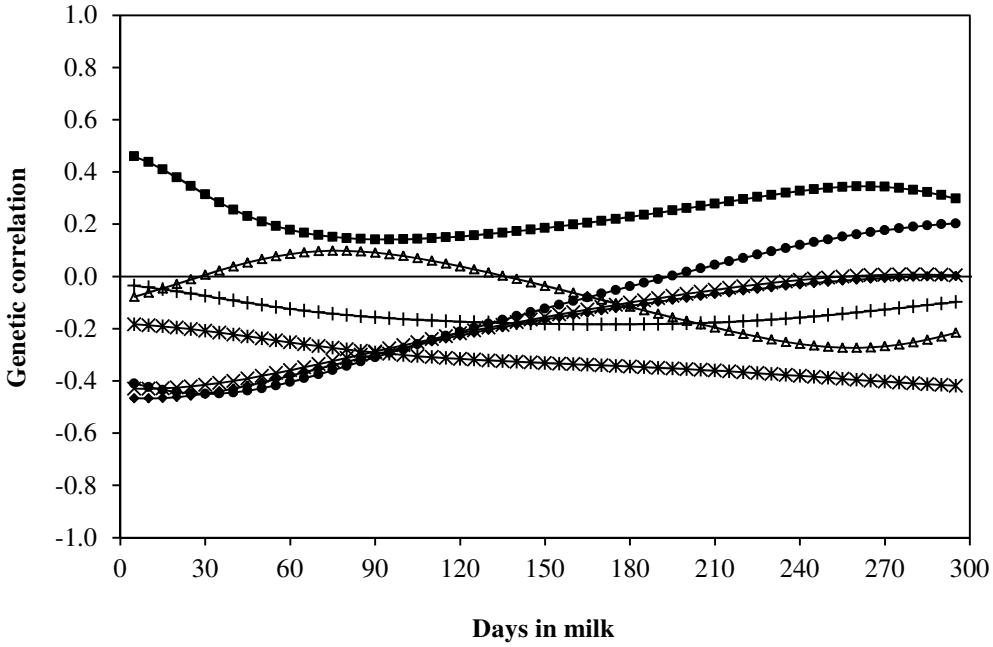


Figure 5.10 Genetic correlations (SE in parenthesis) between â* and b̂* (—♦—; 0.04 to 0.07), milk yield (—Δ—; 0.06 to 0.11), protein concentration (—●—; 0.03 to 0.08), fat concentration (—×—; 0.04 to 0.08), lactose concentration (—*—; 0.04 to 0.10), urea concentration (—+—; 0.05 to 0.11), and $\log_{10}$(SCC/1,000) (—■—; 0.07 to 0.16).

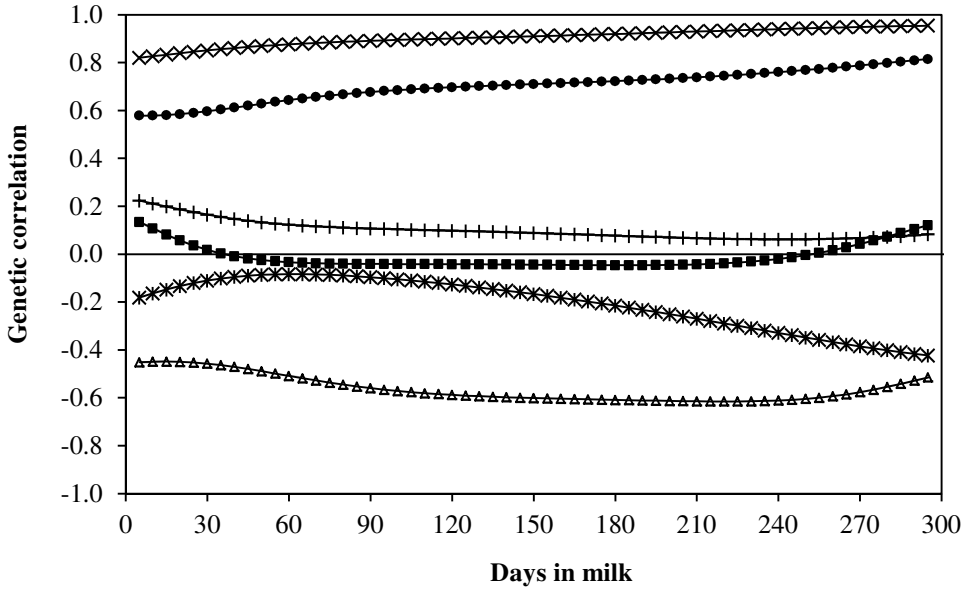


Figure 5.11 Genetic correlations (SE in parenthesis) between b̂* and milk yield (—Δ—; 0.02 to 0.06), protein concentration (—●—; 0.01 to 0.06), fat concentration (—×—; 0.01 to 0.03), lactose concentration (—*—; 0.02 to 0.09), urea concentration (—+—; 0.03 to 0.10), and $\log_{10}$(SCC/1,000) (—■—; 0.04 to 0.18)

The very strong positive genetic correlation averaged across all DIM between yellow colour ($\hat{b}$* index) and milk fat concentration (0.91) mirrored the strong phenotypic correlation obtained both between gold standard (0.64) and predicted (0.77) values. The strong correlation also corroborates previous genetic studies in dairy cows (Winkelman et al., 1999; Morris et al., 2002). The biological justification for such a strong correlation could be due to the presence of β-carotene pigment in milk fat components (MacGibbon et al., 2006; Noziere et al., 2006) which also affect milk colour. Carotenoid pigments are particularly high in fresh grass which was the basal diet of the cows in the present study. The low coefficient of genetic variation of b* parameters after the genetic adjustment for milk fat concentration implies minimal scope to alter milk b* colour genetically independent of genetic merit for milk fat concentration.

5.6 Conclusions

Milking time, stage of lactation, Jersey proportion, parity, and month of test were associated with all three characteristics of milk colour. Simultaneously, heterosis and recombination loss coefficients, as well as the proportions of Montbeliarde, Norwegian Red, and Friesian had little biological impact of the colour of bovine milk. Of particular interest was the potential to breed for different milk colour depending on the respective market demands although the heritability for most of the milk colour traits was not high. The genetic variation was relatively small especially that independent of milk yield and our fat concentration. This therefore suggests that although the accuracy selection is achievable, the ability to rapidly alter milk colour independent of milk yield or fat composition is somewhat limited.

5.7 Acknowledgements

Funding from European Union (Brussels, Belgium) project LowInputBreed and Breed Quality, and laboratory technical support by Bernard Corrigan, Jim Flynn, Paula O'Connor, and Owen Kenny (Teagasc, Moorepark, Fermoy, Co. Cork, Ireland) are gratefully acknowledged.

CHAPTER 6

Potential to breed for improved milk composition of protein fractions and free amino acid concentration in dairy cows

A. McDermott[1,2], S. McParland[1†] M. De Marchi[2], G. Visentin[1,2] and D.P. Berry[1]

[1]Teagasc Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland

[2]Department of Agronomy, Food, Natural Resources, Animals and Environment, University of Padova, Italy

In peer review

6.1 Abstract

Detailed milk composition is not considered in current national dairy cow breeding, despite its known contribution to dairy food product portfolio and yield. Due to the lack of resources required to routinely measure milk compositional traits; there is a deficit in large quantities of phenotypic and genetic data for milk compositional traits. The objective of the present study was to measure the potential of breeding for improved milk protein and free amino acid (FAA) composition, but doing so by exploiting the prediction of these components from routinely available mid-infrared spectroscopy (MIRS) collected from individual cows at milk testing. Milk protein fractions and FAA were measured using available MIRS equations. Genetic, permanent environmental and residual (co) variances for protein fractions and FAA composition were quantified on 134,546 test-day records from 16,166 lactations on 9,572 cows using linear mixed models. Heritability estimates for the gold standard protein fractions ranged from 0.04 (beta casein) to 0.61 (total lactoglobulin) and the range in heritability estimates for the MIRS-predicted protein was less (0.19 for alpha lactalbumin to 0.46 for beta lactoglobulin A). Similar to the protein fractions, heritability estimates for MIRS-predicted FAA had a narrower range (0.15 for glycine to 0.36 for aspartic acid) than the respective gold standard range (0.05 for aspartic acid to 0.58 for serine). The estimated genetic standard deviation for each protein fraction trait genetically independent of protein content was less than the respective unadjusted measure and this was also reflected in lower heritability estimates. There was little impact on the heritability estimates for FAA when adjusted for differences in the genetic merit of 24 hour milk yield. Protein fractions predicted by MIRS were negatively correlated with 24 hour milk yield but positively correlated with protein content and casein content. Genetic correlations among the MIRS-predicted protein fractions were weak to strong. Genetic correlations among the MIRS-predicted FAA were also weak to strong and ranged from -0.44 (aspartic acid and lysine) to 0.97 (glutamic acid and total FAA) and adjusting the correlations for the genetic merit of 24 hour milk yield did not greatly affect the correlations. Results from the current study indicate the presence of exploitable genetic variation for protein fractions and FAA; these traits can be included in the selection index at no marginal cost, because individual cow (and bulk tank) milk samples are routinely subjected to MIRS analysis.

6.2 Introduction

Although most dairy cow breeding objectives include milk protein and fat concentration at the macro level (Miglio et al., 2005), few consider detailed milk composition. This is despite the known contribution of detailed milk composition to dairy food product portfolio and yield (Wedholm et al., 2006; Bonfatti et al., 2011a). For example the concentration of casein (CN) in milk protein has a favourable effect on the quantity of protein transferred from milk into cheese curd. High concentrations of alpha-$_{S1}$-casein (α_{s1}-CN), beta casein (β-CN), kappa casein (κ-CN), and beta lactoglobulin B (β-LG B) are known to increase cheese yield (Wedholm et al., 2006).

Breed differences in the concentration both of protein fractions (Cerbulis et al., 1975; Auldist et al., 2004; Lopez-Villalobos, 2012; **Chapter 4**) and free amino acids (FAA) (**Chapter 4**) have been demonstrated. Individual protein fractions are known to be heritable (Graml and Pirchner, 2003; Schopen et al., 2009; Bonfatti et al., 2011b; Haung et al., 2012) although less is known about the genetic parameters of FAA. Previous heritability estimates of milk protein fractions are moderate to high but differ among studies; recent heritability estimates by Schopen et al. (2009) and Huang et al. (2012) ranged from 0.25 (β-CN) to 0.80 (β-LG) and from 0.33 [α_{S1}-CN, β-CN and alpha lactalbumin (α-LA)] to 0.68 (β-LG), respectively. Differences in the gold standard methods used to quantify milk protein fractions, as well as the characteristics of the study population such as breeds used, and both the parities and stages of lactations represented could have contributed to the difference in estimates (Schopen et al., 2009). Heritability estimates calculated by Bonfatti et al. (2011b) for the relative proportions of protein fractions expressed as a percentage of total CN (0.18 for gamma casein (γ-CN) to 0.69 for β-CN) were greater than those for protein fraction contents expressed as g/L of milk [0.11g/L for α-LA to 0.53g/L for κ-CN]. Although the usefulness of mid-infrared spectroscopy (MIRS) analysis of milk to predict milk composition is now well established (De Marchi et al., 2014) no heritability estimates exist for MIRS-predicted protein fractions or MIRS-predicted FAA.

The objective therefore of the present study was to quantify the potential of breeding for improved milk protein and FAA composition, but doing so by exploiting the prediction of these components from routinely available MIRS collected from individual cows at milk testing.

6.3 Material and methods

6.3.1 Milk sample collection

6.3.1.2 Gold standard data

Milk samples (n=715) were collected from seven research farms operated by the Teagasc Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork between August 2013 and August 2014, inclusive. Individual milk proteins (α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, β-LG A, and β-LG B) were determined for 557 milk samples and FAA [total FAA, glutamic acid (Glu), glycine (Gly), lysine (Lys), arginine (Arg), aspartic acid (Asp), serine (Ser) and valine (Val)] were determined for 715 milk samples using high performance liquid chromatography, as described in detail in **Chapter 2.** Gold standard data were used to calibrate equations to predict individual and groups of milk proteins and FAA using the mid-infrared spectrum of milk (**Chapter 2**).

6.3.1.3 MIRS data

Additional milk samples were collected from seven research herds operated by the Animal and Grassland Research and Innovation Centre, Teagasc, Ireland and from 69 commercial dairy herds located in the south-west of Ireland, between the years 2013 and 2015, inclusive. All milk samples were analysed using the same Fourier transform infrared spectrometer (Foss MilkoScan FT6000 (Foss Electronic A/S, Hillerød, Denmark) based at the Animal and Grassland Research Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland. Prediction equations were developed in **Chapter 2** with untreated spectra using partial least squares regression analysis (Proc PLS; SAS Institute Inc.). The prediction equations were applied to 171,279 spectra from 10,162 cows (17,353 cow lactations) to predict 1) groups of milk proteins (total CN, total whey, total LG), 2) individual proteins (α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, β-LG A, and β-LG B) and 3) FAA (total FAA, Glu, Gly, Lys, Arg, Asp, Ser and Val).

6.3.2 Data editing

Spectral data with a Mahalanobis distance greater than three (Williams 2007) relative to the mean of the 715 gold standard samples were discarded. Furthermore, mid-infrared spectroscopy predicted and gold standard values of proteins and FAA greater than three standard deviations from the mean of the gold standard samples were also removed from the analyses. Traits that were not normally distributed (i.e. Glu, Gly, Lys,

Arg, Asp, Ser and Val) were transformed using a natural logarithm transformation. Breed composition was available for each cow and only data from Holstein, Friesian, Jersey, Norwegian Red, Montbelliarde cows as well as their crosses recorded between 5 and 305 DIM and from parities ≤10 were retained for analysis; parties greater than 5 were grouped together for analysis.

Contemporary groups were generated for the research animals according to experimental treatment and test-date; contemporary groups for commercial animals were defined as herd-test-date. For the MIRS data only, contemporary groups with a minimum of ten records were retained for analysis. The final MIRS data set comprised 134,546 records from 16,166 lactations on 9,572 cows.

6.3.3 Data analysis

Pedigree information for all animals was provided by the Irish Cattle Breeding Federation database and each animal was traced back (where available) at least four generations. The pedigree file contained 33,949 animals. Genetic, permanent environmental and residual (co) variances for protein fractions and FAA composition were quantified using linear mixed models in ASReml (Gilmour et al., 2009). Models were adjusted for the fixed effects of contemporary group, parity (1, 2, 3, 4, 5 and ≥ 6), stage of lactation (6 groups each 60 day in length from 5 DIM to 305 DIM), the interaction between parity and stage of lactation, milking time (i.e., AM or PM), proportion of cow breed (Friesian, Jersey, Norwegian Red, Montbelliarde and other), general heterosis and recombination loss coefficients of the cow. Random effects included the direct additive genetic effect of the animal and both a within- and an across-lactation cow permanent environmental effect. Models to analyse the gold standard traits did not include contemporary group or cow permanent environmental effects but no gold standard repeated records existed. Genetic correlations between the same gold standard and MIRS-predicted traits were estimated using the aforementioned model.

Genetic and phenotypic (co)variances among the MIRS-predicted protein fractions were estimated using a series of trivariate analyses which, as well as including the two MIRS-predicted protein fractions, also included total protein content. Genetic and phenotypic (co)variances among the MIRS-predicted FAA were estimated using a series of trivariate analyses which as well as including the two MIRS-predicted FAA, also included 24 hour milk yield. Genetic correlations among the MIRS-predicted protein fractions adjusted for their respective genetic correlation with total protein content as well

as the genetic correlation among the MIRS-predicted FAA adjusted for their respective genetic correlation with 24 hour milk yield were estimated as:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

where $r_{xy.z}$ is the partial correlation between trait x and y independent of z, r_{xy} represents the correlation between the traits x and y, r_{xz} is the correlation between x and z and r_{yz} is the correlation between y and z. The phenotypic and genetic variances for the protein fractions and the FAA adjusted for their respective genetic and phenotypic correlation with protein content or milk yield was calculated as

$$\sigma^2 * (1 - r_{xy}^2)$$

Where σ^2 is the variance for the protein fraction or the FAA and r_{xy} is the correlation between trait x (protein fraction or FAA) and trait y (protein content and milk yield).

6.4 Results

The mean values of the gold standard and the respective MIRS-predicted milk protein fractions were similar; for example the mean value of the gold standard total CN and the mean MIRS-predicted total CN was 36.91 g/L and 36.16 g/L, respectively (**Table 6.1 and 6.2**). The mean values of the gold standard and the respective MIRS-predicted milk FAA variables were also similar to each other. The genetic standard deviation ranged from 0.16 g/L (β-LG A) to 1.36 g/L (total CN) for the gold standard protein fractions (**Table 6.1**) and from 0.18 g/L (α_{S2}-CN) to 1.85 g/L (total CN) for the MIRS-predicted protein fractions (**Table 6.2**). The genetic standard deviation of the gold standard and the respective MIRS-predicted FAA variables were similar, with the exception of total FAA (13.35 µg/ml for the gold standard but 5.05 µg/ml for MIRS-prediction). The coefficient of genetic variation differed among traits and ranged from 3.01 (total LG) to 43.11 (β-LG B) for the gold standard protein fractions and from 1.01 (Glu) to 25.26 (Ser) for the gold standard FAA (**Table 6.1**).

Table 6.1 Number of records (n), mean, genetic standard deviation (σg), heritability (SE), coefficient of genetic variation (CVg) for gold standard protein fractions and gold standard free amino acids and as well as the genetic correlation (rg) between gold standard and mid-infrared spectroscopy predicted protein fractions and gold standard and mid-infrared spectroscopy predicted free amino acids.

Trait	n	Mean	σ_g	Heritability	CV_g	r_g
Protein, g/L						
Total Casein	417	36.91	1.36	0.17 (0.160)	3.68	0.74 (0.18)
Alpha $_{S1}$ Casein	415	14.22	0.94	0.39 (0.173)	6.58	0.63 (0.12)
Alpha $_{S2}$ Casein	415	3.71	0.50	0.49 (0.166)	13.37	0.36 (0.11)
Beta Casein	415	12.99	0.54	0.04 (0.148)	4.13	0.99 (1.84)
Kappa Casein	416	6.14	0.34	0.06 (0.149)	5.54	0.43 (0.12)
Total Whey	418	6.33	1.14	0.57 (0.194)	18.00	0.74 (0.12)
Alpha Lactalbumin	412	1.13	0.03	0.35 (0.195)	3.01	0.64 (0.17)
Total Lactoglobulin	418	5.12	1.18	0.61 (0.193)	22.98	0.76 (0.11)
Beta Lactoglobulin A	417	2.58	0.16	0.19 (0.143)	6.22	0.68 (0.17)
Beta Lactoglobulin B	383	2.79	1.20	0.55 (0.189)	43.11	0.82 (0.06)
Free AA, µg/mL						
Total free AA [1]	463	52.85	13.35	0.37 (0.154)	25.27	0.53 (0.12)
Glutamic Acid[1]	461	30.57	0.31	0.29 (0.148)	1.01	0.78 (0.13)
Glycine[1]	453	7.23	0.26	0.32 (0.154)	3.64	0.68 (0.18)
Lysine[1]	441	4.66	0.18	0.25 (0.158)	3.85	0.34 (0.26)
Arginine[1]	419	4.20	0.30	0.38 (0.182)	7.14	0.96 (8.62)
Aspartic Acid[1]	385	2.65	0.15	0.05 (0.138)	5.82	0.94 (0.29)
Serine[1]	389	1.49	0.38	0.58 (0.188)	25.65	0.38 (0.11)
Valine[1]	424	1.98	0.30	0.28 (0.157)	15.13	0.46 (0.17)

[1]Traits were log-transformed before analysis

Table 6.2 Number of records (n), mean, genetic standard deviation (σ_g), heritability (standard error) and repeatability (standard error), as well as heritability estimates after adjustment for protein content or 24 hour milk yield for MIRS protein fractions and MIRS free amino acids.

			Unadjusted			After Adjustment	
Trait	n	Mean	σ_g	Heritability	Repeatability	σ_g	Heritability
Protein, g/L[1]							
Total casein	134,395	36.16	1.85	0.42 (0.017)	0.55 (0.004)	0.27	0.11
Alpha $_{S1}$ casein	134,409	13.70	0.79	0.44 (0.017)	0.41 (0.006)	0.12	0.13
Alpha $_{S2}$ casein	134,482	3.66	0.18	0.36 (0.015)	0.44 (0.006)	0.06	0.11
Beta casein	134,235	12.96	0.54	0.38 (0.016)	0.32 (0.006)	0.18	0.16
Kappa casein	134,463	6.03	0.34	0.36 (0.016)	0.39 (0.006)	0.09	0.16
Total whey	134,383	6.13	0.50	0.37 (0.015)	0.34 (0.006)	0.46	0.35
Alpha lactalbumin	134,106	1.11	0.04	0.19 (0.011)	0.44 (0.006)	0.03	0.15
Total lactoglobulin	134,236	5.09	0.51	0.39 (0.016)	0.22 (0.005)	0.44	0.37
Beta lactoglobulin A	134,489	2.39	0.19	0.46 (0.016)	0.46 (0.006)	0.10	0.31
Beta lactoglobulin B	128,391	2.61	0.56	0.43 (0.017)	0.48 (0.006)	0.55	0.43
Free AA, µg/mL[2]							
Total free AA	134,546	53.60	5.05	0.24 (0.096)	0.36 (0.006)	5.05	0.24
Glutamic acid	134,426	30.93	0.14	0.32 (0.016)	0.44 (0.006)	0.14	0.32
Glycine	133,650	8.09	0.08	0.15 (0.011)	0.23 (0.005)	0.08	0.15
Lysine	134,105	4.75	0.12	0.24 (0.013)	0.31 (0.006)	0.11	0.22
Arginine	134,425	3.38	0.07	0.19 (0.012)	0.28 (0.005)	0.06	0.19
Aspartic acid	134,433	2.73	0.15	0.36 (0.016)	0.47 (0.006)	0.15	0.36
Serine	112,918	2.74	0.10	0.23 (0.013)	0.31 (0.006)	0.10	0.22
Valine	133,957	1.52	0.09	0.24 (0.014)	0.30 (0.006)	0.09	0.18

[1]Adjusted for protein content.

[2]Adjusted for 24 hour milk yield and log-transformed before analysis.

6.4.1 Heritability and repeatability estimates

Heritability estimates for the gold standard protein fractions, albeit associated with large standard errors, ranged from 0.04 (β-CN) to 0.61 (total LG) (**Table 6.1**). Heritability estimates for MIRS-predicted protein fractions were higher than those estimated for the gold standard protein fractions (**Table 6.2**), although the range in heritability estimates for the MIRS-predicted protein fractions was less (0.19 for α-LA to 0.46 for β-LG A; **Table 6.2**) than the range for the respective gold standard measures (0.04 for β-CN to 0.61 for total Lg; **Table 6.1**). Similar to the protein fractions, heritability estimates for MIRS-predicted FAA had a narrower range (0.15 for Gly to 0.36 for Asp; **Table 6.2**) than the respective gold standard range (0.05 for Asp to 0.58 for Ser; **Table 6.1**). Repeatability estimates for MIRS-predicted protein fractions and MIRS-predicted FAA were moderate (0.22 for total LG to 0.55 for total CN; 0.23 for Gly to 0.47 for Asp; **Table 6.2**). The estimated genetic standard deviation for each protein fraction trait genetically independent of protein content was less than the respective unadjusted measure and this was also reflected in lower heritability estimates; only a small decrease in genetic standard deviation and heritability was observed for the whey fractions. There was also little impact on the heritability estimates for FAA when adjusted for differences in the genetic merit of 24 hour milk yield.

6.4.2 Genetic correlations

Moderate to strong genetic correlations existed between the gold standard protein fractions and their respective MIRS-predicted protein fractions, ranging from 0.36 (β-CN) to 0.99 (κ-CN) (**Table 6.1**). The genetic correlation between the gold standard Arg and MIRS-predicted Arg was strong (0.96), similar to the correlation between the gold standard and MIRS-predicted Asp (0.94). Protein fractions were negatively correlated with 24 hour milk yield but positively correlated with protein content and CN content (**Table 6.3**). Protein fractions were weakly correlated with both lactose content and somatic cell count (SCC). The protein fractions α_{S1}-CN, κ-CN, total whey, total LG, and β-LG B were negatively correlated with SCC. Individual FAA were weakly to moderately genetically correlated with all of 24 hour milk yield, protein content, CN content, fat content, lactose content, and SCC (**Table 6.4**). Correlations between FAA and SCC were weak while total FAA, Glu, Gly, Asp, and Ser were all negatively correlated with SCC.

Genetic correlations among the MIRS-predicted protein fractions were weak to strong (**Table 6.5**). An almost unity genetic correlation (0.99) existed between protein content and total CN, between total CN and α_{S1}-CN, and between total-LG and total whey. The genetic correlation between total-LG and total whey did not change after adjusting for their respective genetic correlation with protein content, albeit the majority of the correlations among the protein fractions weakened when calculated genetically independent of protein content. For example, the strong positive genetic correlation that existed between β-CN and β-LG B (0.89) became negative (-0.05) once adjusted for genetic merit for protein content.

Genetic correlations among the MIRS-predicted FAA were weak to strong and ranged from -0.44 (Asp and Lys) to 0.97 (Glu and Total FAA) (**Table 6.6**). Adjusting the correlations for the genetic merit of 24 hour milk yield did not greatly affect the correlations (**Table 6.6**). For example, the unadjusted correlation between Arg and Val was 0.63 and changed to 0.62 when adjusted for the genetic merit of 24 hour milk yield.

6.5 Discussion

The present study aimed to quantify the extent of genetic variability in detailed milk protein and FAA composition predicted by MIRS from individual cow milk samples. Global milk recording programmes are presently using MIRS to determine the concentration of fat and protein in milk samples for herd testing and genetic evaluations. Protein composition and FAA composition are not, however, routinely determined by MIRS or used in national genetic evaluations. The genetic correlations estimated in the present study between both the gold standard and MIRS-predicted protein fractions, as well as between both the gold standard and MIRS-predicted FAA, were moderate to strong demonstrating that the MIRS-predicted traits are genetically very similar to their corresponding gold standard measures. All traits were heritable and exhibited considerable genetic variation; therefore MIRS could be a viable method to collect a large amount of data for use in genetic evaluations with the goal of improving milk quality traits.

Table 6.3 Genetic correlations between protein fractions and 24 h milk yield, MIRS-predicted protein content, MIRS-predicted casein content, MIRS-predicted fat content, MIRS-predicted lactose content and somatic cell count (SCC).

	24 hour Milk yield	Protein content	Casein content	Fat content	Lactose content	SCC
Total-CN	-0.57(0.040)	0.99(0.001)	0.99(0.001)	0.72(0.021)	-0.09(0.038)	0.01(0.105)
α_{S1}-CN	-0.58(0.039)	0.99(0.001)	0.98(0.002)	0.74(0.019)	-0.12(0.037)	-0.01(0.105)
α_{S2}-CN	-0.59(0.040)	0.95(0.005)	0.96(0.004)	0.79(0.018)	-0.05(0.038)	0.01(0.106)
β-CN	-0.57(0.042)	0.94(0.005)	0.81(0.032)	0.69(0.023)	0.06(0.039)	0.02(0.107)
k-CN	-0.50(0.043)	0.97(0.003)	0.94(0.005)	0.62(0.025)	-0.20(0.038)	-0.02(0.109)
Total-Whey	-0.30(0.048)	0.49(0.027)	0.46(0.029)	0.39(0.031)	-0.07(0.038)	-0.09(0.110)
α-LA	-0.45(0.052)	0.55(0.031)	0.59(0.030)	0.68(0.026)	0.35(0.038)	0.01(0.117)
Total-Lg	-0.29(0.048)	0.49(0.027)	0.46(0.029)	0.38(0.031)	-0.12(0.037)	-0.10(0.110)
β-Lg-A	-0.56(0.040)	0.86(0.010)	0.87(0.009)	0.75(0.018)	-0.09(0.036)	0.04(0.103)
β-Lg-B	-0.08(0.051)	0.18(0.034)	0.15(0.035)	0.11(0.036)	-0.11(0.038)	-0.12(0.113)

[1] Total-Casein (Total-CN), Alpha-$_{S1}$-Casein (α_{S1}-CN), Alpha-$_{S2}$-Casein (α_{S1}-CN), Beta Casein (β-Casein), Kappa Casein (k-CN), Alpha-Lactalbumin (α-LA), Total Lactoglobulin (Total-LG) and Beta-Lactoglobulin A (β-LG A), Beta-Lactoglobulin B (β-LG B).

Table 6.4 Genetic correlations between FAA and 24 hour milk yield, MIRS-predicted protein content, MIRS-predicted casein content, MIRS-predicted fat content, MIRS-predicted lactose content and somatic cell count (SCC).

	24 hour Milk yield	Protein content	Casein content	Fat content	Lactose content	SCC
TFAA[1]	0.01(0.059)	-0.10(0.041)	-0.09(0.041)	-0.07(0.042)	0.29(0.042)	-0.11(0.120)
Glu[1]	0.10(0.054)	-0.18(0.037)	-0.17(0.038)	-0.15(0.039)	0.33(0.037)	-0.04(0.112)
Gly[1]	0.05(0.063)	-0.19(0.043)	-0.17(0.044)	-0.24(0.043)	0.28(0.046)	-0.24(0.123)
Lys[1]	-0.33(0.053)	0.52(0.030)	0.51(0.031)	0.51(0.032)	-0.16(0.041)	0.04(0.114)
Arg[1]	-0.18(0.059)	0.21(0.041)	0.17(0.042)	0.40(0.036)	-0.33(0.041)	0.13(0.119)
Asp[1]	0.14(0.053)	-0.19(0.036)	-0.20(0.036)	-0.16(0.038)	0.25(0.038)	-0.08(0.112)
Ser[1]	-0.13(0.058)	-0.15(0.039)	-0.11(0.040)	0.24(0.039)	0.26(0.041)	-0.04(0.118)
Val[1]	-0.21(0.060)	0.32(0.039)	0.33(0.040)	0.25(0.042)	-0.16(0.041)	0.09(0.121)

[1]Traits were log-transformed before analysis.

Table 6.5 Genetic correlations among MIRS-predicted protein fractions (above diagonal) and genetic correlations among MIRS-predicted protein fractions genetically independent of protein content (below the diagonal).

	Total CN	α_{S1}-CN	α_{S2}-CN	β-CN	k-CN	Total whey	α-LA	Total-LG	β-LG A	β-Lg B
Total CN	-	0.99(0.001)	0.92(0.005)	0.97(0.003)	0.96(0.003)	0.47(0.028)	0.60(0.028)	0.47(0.028)	0.85(0.009)	0.17(0.034)
α_{S1}-CN	0.91	-	0.92(0.006)	0.96(0.003)	0.95(0.004)	0.51(0.026)	0.61(0.028)	0.50(0.026)	0.81(0.012)	0.18(0.033)
α_{S2}-CN	-0.24	-0.24	-	0.89(0.008)	0.90(0.007)	0.42(0.029)	0.56(0.029)	0.41(0.029)	0.88(0.007)	0.10(0.034)
β-CN	0.79	0.63	-0.05	-	0.80(0.016)	0.48(0.028)	0.70(0.024)	0.46(0.029)	0.86(0.010)	0.13(0.035)
k-CN	0.07	-0.09	-0.12	-0.22	-	0.47(0.028)	0.38(0.037)	0.47(0.028)	0.72(0.017)	0.22(0.034)
Total whey	0.09	0.16	-0.17	0.05	-0.01	-	0.47(0.032)	0.99(0.001)	0.35(0.029)	0.91(0.007)
α-LA	0.53	0.56	0.14	0.63	-0.37	0.25	-	0.44(0.033)	0.72(0.022)	0.17(0.039)
Total-LG	0.04	0.12	-0.17	-0.02	0.01	0.99	0.24	-	0.34(0.030)	0.92(0.006)
β-Lg A	0.11	0.25	0.44	0.29	-0.79	-0.01	0.58	-0.18	-	-0.03(0.033)
β-Lg B	-0.05	-0.01	-0.20	-0.13	0.18	0.96	0.09	0.90	-0.36	-

[1] Total Casein (Total-CN), Alpha-$_{S1}$-casein (α_{S1}-CN), Alpha-$_{S2}$-casein (α_{S1}-CN), Beta casein (β-Casein), Kappa casein (k-CN), Alpha lactalbumin (α-LA), Total lactoglobulin (Total-LG) and Beta lactoglobulin A (β-LG A), Beta lactoglobulin B (β-LG B).

Table 6.6 Genetic correlations among MIRS-predicted FAA (above diagonal) and genetic correlations among MIRS-predicted FAA genetically independent of 24 hour milk yield (below the diagonal).

Traits	TFAA	Glu	Gly	Lys	Arg	Asp	Ser	Val
TFAA[1]	-	0.97(0.003)	0.56(0.035)	-0.20(0.044)	0.31(0.043)	0.88(0.011)	0.60(0.029)	0.49(0.036)
Glu[1]	0.99	-	0.53(0.040)	-0.35(0.038)	0.19(0.040)	0.91(0.008)	0.57(0.030)	0.42(0.037)
Gly[1]	0.56	0.53	-	-0.36(0.044)	-0.27(0.050)	0.57(0.033)	0.37(0.040)	-0.09(0.053)
Lys[1]	-0.20	-0.31	-0.36	-	0.56(0.032)	-0.44(0.035)	0.06(0.044)	0.49(0.037)
Arg[1]	0.24	0.21	-0.22	0.53	-	0.13(0.040)	0.32(0.041)	0.63(0.030)
Asp[1]	0.90	0.90	0.57	-0.42	0.16	-	0.32(0.040)	0.18(0.043)
Ser[1]	0.38	0.59	0.39	0.02	0.30	0.30	-	0.49(0.037)
Val[1]	0.49	0.45	-0.09	0.45	0.62	0.19	0.27	-

[1]Traits were log-transformed before analysis.

6.5.1 Milk protein fractions and free amino acids response to genetic selection

Response to genetic selection is determined by the extent of genetic variability, the accuracy of selection, selection intensity and generation interval (Rendel and Robertson, 1950). Therefore to achieve genetic gain, the trait must be heritable, exhibit genetic variation and sufficient data on the trait must be available to ensure a high accuracy of selection. Considerable genetic variation clearly exists for all milk quality traits examined in the present study and traits were low to moderately heritable; this therefore suggests that there is potential to alter the protein composition and FAA composition in bovine milk using selective breeding, and high accuracy of selection could be achieved for sires based on relatively small progeny group sizes.

Heritability estimates of all the MIRS-predicted protein fractions in the present study (0.19 to 0.46; **Table 6.2**) were higher than the heritability estimate for 24 hour milk yield (0.17). Moreover, the heritability of 24 hour milk yield (0.17) and protein content (0.46) in the present study were in the range of the heritability estimates for the gold standard protein fractions (0.04 to 0.61; **Table 6.1**), the gold standard FAA (0.05 to 0.58; **Table 6.1**) and the MIRS-predicted FAA (0.15 to 0.36; **Table 6.2**). Previous heritability estimates of protein fractions in bovine milk have been documented to be moderate to high ranging from 0.25 for β-CN to 0.80 for total LG (Schopen et al., 2009, Bonfatti et al., 2011; Haung et al., 2012), although they differed across studies. Recent heritability estimates for gold standard protein fractions in milk from Simmental cows ranged from 0.18 (κ-CN) to 0.68 (α_{S1}-CN) (Bonfatti et al., 2009) and in milk from both Holstein and Holstein-Jersey crosses the range was between 0.33 [α_{S1}-CN, β-CN and α-LA) and 0.68 (β-LG) (Huang et al., 2012).

In the present study, all protein fractions were positively genetically correlated with protein content contradicting previous findings by Schopen et al. (2009), who reported β-CN, κ-CN, α-LA, β-LG, and total whey to be negatively genetically correlated with protein content. Genetic correlations among the protein fractions were also found to be different between the present study and between the studies of Schopen et al. (2009) and Bonfatti et al. (2011). However, different methods used to measure milk protein fractions, as well as the characteristics of the study populations such as the breeds used, as well as the parities and stages of lactations represented, could have contributed to the difference in correlations. Bonfatti et al. (2011) used milk samples from 2,167 Simmental

cows and protein fractions were determined by high performance liquid chromatography, whereas Schopen et al. (2009) used information from 1,940 first-parity Holstein Friesian cows and protein fractions were determined by capillary zone electrophoresis. In the present study, protein fractions were predicted by MIRS for up to 134,100 milk samples, from Holstein, Friesian, Jersey, Norwegian Red, and Montbelliarde cows as well as their crosses, from a range of parities and stages of lactations.

Many current national breeding programmes indirectly select for protein content through a negative weighting on milk yield concurrent with a positive weighting on protein yield. The coefficient of genetic variation for milk yield and protein content in the present study was 7.45 and 4.80, respectively; these were less than the average coefficient of genetic variation for the gold standard protein fractions (12.62) and FAA (10.94). Genetic gain in the milk yield of dairy cows in recent years is well documented (Berry *et al.,* 2014; Berry, 2008; Norman and Powell, 1999) with lactation milk yield per cow doubling over the past 40 years (Oltenacu and Broom, 2010); given the respective coefficient of variation for milk yield, protein fractions and FAA this therefore implies that similar genetic gain is possible for protein fractions and FAA if these traits were included in a selection index with high accuracy of selection.

However, a relevant question is the benefit in response to selection of including these detailed proteins or FAA in a breeding objective which already includes other traits like milk yield and protein content. One approach to estimate such a benefit is the coefficient of genetic variation of these traits after adjustment for the genetic merit of protein content or milk yield. The near unity correlation between milk protein content and total CN (0.99) and the resulting small coefficient of genetic variation for total CN after adjustment for the genetic merit of protein content (0.75) in the present study suggests that including CN content directly in a breeding goal may be of little additional benefit. Similar to the correlation between milk protein content and total CN, the correlations between α_{S1}-CN and protein content, between β-CN and protein content and between κ-CN and protein content were strong (0.95, 0.94 and 0.97, respectively) and the coefficient of genetic variation for α_{S1}-CN, β-CN and κ-CN after adjustment for the genetic merit of protein content was 0.88, 1.39 and 1.49, respectively. Hence, a high selection pressure would need to be applied to α_{S1}-CN, β-CN and κ-CN and this would result in less selection pressure on the other traits within the breeding index and thus reduced genetic gain. The correlation between protein content and α-LA was only 0.55 and the

corresponding coefficient of genetic variation after adjusting for the genetic merit of protein content was 2.70. Therefore, it may advantageous to include α-LA as an individual trait in the breeding index even if genetic gain is slow, as a higher concentration of α-LA in milk is desirable in infant formula production (Lien, 2003).

The weak genetic correlations between FAA and 24 hour milk yield and between FAA and protein content signify that current selection objectives on milk yield and protein content (Miglior et al., 2005) are not fully exploiting the potential to genetically improve milk protein fraction and FAA phenotypes. The near unity genetic correlation between total FAA and Glu (0.99) implies there is little expected benefit of including both in a breeding index. The lack of very strong genetic correlations among the other FAA (i.e., Gly, Lys, Arg, Asp, Ser and Val) plus the existence of a coefficient of genetic variation for these traits ranging from 3.64 (Gly) to 25.65 (Ser; **Table 6.1**) signify these traits could be simultaneously in a breeding index.

6.5.2 Practical implications of results

Routine access to vast quantities of phenotypic data for protein fractions and FAA is imperative to achieve a high accuracy of selection and thus genetic gain. Based on the heritability estimates, as well as the phenotypic and genetic variances of the traits estimated in the present study, the number of progeny required to achieve a reliability (i.e., squared accuracy of selection) of 0.70 for a sire is 20 for α_{S1}-CN, 47 for α-LA, 170 for Glu, and 65 for Gly. MIRS is an efficient method commonly used by milk recording organisations worldwide to predict milk fat, protein, CN and lactose and the ability of MIRS to predict individual protein fractions and FAA with reasonable accuracies has been previously documented (**Chapter 2**). Thus, MIRS could be used as a rapid and cost-effective method for generating substantial quantities of milk protein fraction and FAA phenotypes; hence, a high accuracy of selection for some protein fractions and FAA is possible.

Of potential interest to many processors are the similarities, or lack thereof, between human and bovine milk and the potential strategies to make them more similar but doing so at a low cost. Human and bovine milk differ in their protein profile; human milk has a whey to casein ratio of approximately 60:40, while bovine milk has a whey to casein ratio of approximately 20:80. Human milk does not contain β-LG, albeit it is present in the greatest amount in bovine milk and the concentration of α-LA (the dominant protein in human milk) is relatively low in bovine milk (Lien, 2003). It is

therefore advantageous for infant formula producers to select bovine milk with a higher concentration of α-LA and a lower concentration of β-LG and being able to breed cows for such a profile would therefore be very beneficial. The amino acid profile in human and bovine milk is also different (Chuang et al., 2005); human milk contains more Glu, Lys, Arg, Asp, Ser and Val than bovine milk (Ghadimi and Pecora, 1963). Indeed, the heritability estimate of the gold standard α-LA was 0.35, while the corresponding coefficient of genetic variation was 3.01 and similarly, the heritability estimate of total FAA was 0.37, with a coefficient of genetic variation of 25.27; therefore, potential clearly exists to breed for both α-LA and total FAA.

Processors also aim to maximise the efficiency of transformation of the milk they purchase into saleable products. A high concentration of CN to total protein is imperative for the transfer of proteins from milk to cheese and high concentrations of α_{S1}-CN, β-CN, κ-CN, and β-LG B increase cheese yield (Wedholm et al., 2006). The heritability estimates of the MIRS-predicted CN fractions ranged from 0.36 (κ-CN and α_{S2}-CN) to 0.44 (α_{S1}-CN) and the genetic standard deviation calculated for gold standard CN fractions in the present study (1.10 g/L for α_{S1}-CN) were comparable to results obtained by Schopen et al. (2009) (0.94 g/L for α_{S1}-CN). Genetic variation for CN fractions also existed in the present study (genetic standard deviation ranged from 0.18 g/L for α_{S2}-CN to 0.79 g/L for α_{S1}-CN), even independent of the genetic merit of protein content indicating ample opportunity to improve the composition of protein fractions in cow milk through breeding, thereby improving the quality of milk for cheese production.

6.6 Conclusions

In conclusion, all traits were heritable and exhibited considerable exploitable genetic variation; therefore MIRS could be a viable method to collect a large amount of data for use in genetic evaluations with the goal of improving milk quality traits. Protein fractions and FAA can be included in the selection index at no marginal cost, because individual cow (and bulk tank) milk samples are routinely subjected to MIRS analysis.

6.7 Acknowledgements

Funding for this work was received from the Irish Department of Agriculture, Food and the Marine, Research Stimulus Fund project 11/SF/311, Breed Quality.

CHAPTER 7

Summary, Overall Conclusions and Implications and Future Research

7.1 Summary

The overall aim of this thesis was to determine the feasibility of breeding for improved milk quality and in particular protein fractions, free amino acids (FAA) and milk colour traits. To breed for a characteristic such as milk quality it must be: (i) economically or socially important (ii) exhibit genetic variation (i.e. be heritable), and (iii) be measurable or genetically correlated with a measurable trait.

- In **Chapters 2** and **3**, milk samples were collected from 7 research farms operated by the Teagasc Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork between August 2013 and August 2014, inclusive. Milk protein fractions and FAA were determined by high performance liquid chromatography; milk colour was measured using a Chroma Meter CR400 with a closed cone, set on the L* a* b* system. Prediction equations were developed to predict 557 proteins, 715 FAA and 601 milk colour samples directly from the MIR spectrum. **Chapters 4** and **6** utilised additional spectral data from two sources: (i) 7 research herds operated by the Animal and Grassland Research and Innovation Centre (Teagasc, Ireland) and (ii) 69 Irish commercial dairy herds. The research spectra comprised of 94,286 separate morning and evening milk samples; the commercial spectra comprised of 40,260 milk samples (morning and evening milk samples combined). Gold standard data (715 milk samples) were used to generate equations to predict individual and groups of milk proteins and FAA using the mid-infrared spectrum of milk. In **Chapter 4** factors associated with both protein and FAA composition traits were quantified separately using linear mixed models in ASReml. In **Chapter 5**, the spectral data consisted of 136,807 milk samples from 9,824 cows and the pedigree of all animals was traced back at least four generations. Mid-infrared spectroscopy prediction models previously developed in **Chapter 2** were applied to all spectra to predict milk lightness (L*), red-green index (a*), and yellow-blue index (b*). Factors associated with milk colour traits were quantified separately using linear mixed models in ASReml. In **Chapter 6**, pedigree information for 33,949 animals was used to quantify the extent of genetic variation in the milk quality traits of MIRS predicted protein fractions and FAA. Genetic, permanent environmental and residual (co) variances for protein fractions and FAA composition were quantified using linear mixed models.

Results from this thesis clearly show that MIRS is useful to routinely and efficiently predict some milk quality traits such as some protein fractions, some FAA and

the b* colour of milk at a population level. The greatest correlation coefficient of external validation obtained for protein fractions, FAA and colour were 0.74 (total CN), 0.75 (Gly) and 0.72 (b*), respectively. The knowledge generated in this thesis of how milk protein fractions and FAA change across calendar months of the year, stage of lactation, parity and breed could provide useful input parameters for decision support tools in the management of product portfolios by processors. For example, protein fractions were present in the lowest concentration in July and August, while FAA were present in the greatest concentration during these months. Younger animals produced more total CN, total whey and total β-LG and some FAA in milk across lactation than their older contemporaries. Jersey cows produced milk that had a greater concentration of all CN fractions but a lesser concentration of total FAA than Holstein cows.

All protein fractions, FAA and colour were low to moderately heritable. Heritability of the predicted protein fractions, FAA and milk colour ranged from 0.04 (β-CN) to 0.61 (total-LG), from 0.05 (Asp) to 0.58 (Ser), and from 0.29 (L*) to 0.35 (b*), respectively. The coefficient of genetic variation of protein fractions, FAA and milk colour ranged from 3.60 (α-LA) to 21.46 (β-LG B), from 0.45 (Glu) to 9.42 (total FAA), and from 0.37 (L*) to 1.72 (a*), respectively.

Results from this thesis demonstrated that some protein fractions, some FAA and milk colour are predictable from MIR and these predictions exhibit genetic variation and thus breeding for improved milk quality is feasible. The prediction of protein fractions, FAA and milk colour by MIRS could benefit the dairy breeding industry worldwide through genetic selection of animals for superior quality milk and allowing for the more accurate selection of milk for human consumption, infant milk formula and cheese production. The generated predictions could also be useful for optimising both herd and processor management strategies.

7.1.1 Chapter 1: Literature review

Objective: To review the available literature pertinent to the mid-infrared spectroscopy prediction of protein fractions, FAA and milk colour as well as the phenotypic and genetic factors associated with these traits.

- Detailed milk product quality traits are not considered in the Irish national dairy cow-breeding objective, at present, despite their importance to the bio-economy.
- Bovine milk generally consists of about 3.3% protein, of which 78% is comprised of CN, 17-18% of whey protein and the remaining 4-5% as non-protein nitrogen.
- Free amino acids (FAA) in milk are amino acids resulting from milk protein denaturation and therefore do not contribute to the total protein of milk.
- Human and bovine milk have different FAA content and composition, with bovine milk generally having a lesser concentration of FAA than human milk.
- The yellow colour of bovine milk is related to the level of β-carotene and fat content in milk with a greater β-carotene content associated with a more yellow colour in milk.
- Mid-infrared spectroscopy (MIRS) is a technique that studies the interactions between light and matter within the mid-infrared region of the electromagnetic spectrum.
- MIR prediction accuracies for CN fractions ranged in value from a coefficient of determination of 0.13 for β-CN to a coefficient of determination of 0.66 for α_{S1}-CN.
- Both genetic and management factors such as breed, parity, stage of lactation and milking time influence the quantity of milk protein fractions, FAA and colour of bovine milk.
- Heritability estimates of individual milk proteins are moderate to high; ranging from 0.33 for α_{S1}-CN, β-CN, α-LA and to 0.69 for β-LG (Haung et al., 2012)
- Genetic correlations among the CN fractions and between CN and whey fractions were generally weak to moderate.
- Gaps in the knowledge include:
    - Effectiveness of MIRS in predicting milk protein fractions, FAA and colour.
    - Cow level factors associated with protein fractions, FAA and colour in milk from grazing dairy cows.
    - Genetic parameters for protein fractions, FAA and colour in milk from grazing dairy cows.

7.1.2 Chapter 2: Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics

Objective: To evaluate the effectiveness of mid-infrared spectroscopy in predicting milk protein fractions and FAA composition in bovine milk.

- Milk protein fractions and FAA were determined by high performance liquid chromatography for 557 and 715 milk samples, respectively.
- Prediction models from MIRS were developed to predict each trait separately using partial least squares regression.
- The correlation coefficient of cross validation between gold standard and MIRS predicted protein fractions ranged from 0.43 (β-LG A) to 0.76 (total whey and LG); the greatest correlation coefficient of external validation value obtained for protein fractions was 0.74 for total CN, respectively.
- The correlation coefficient of cross validation between gold standard and MIRS predicted FAA ranged from 0.51 (Ser) to 0.75 (Gly); the greatest correlation coefficient of external validation value obtained for FAA was 0.75 for Gly.
- The linear regression coefficient of the gold standard values on the respective MIRS protein fractions ranged from 0.76 (β-LG B) to 0.99 (β-CN and κ-CN) and the bias ranged from -0.0068 g/L (total CN) to 0.0072 g/L (α_{S2}-CN).
- The linear regression coefficient of the gold standard values on the respective MIRS protein fractions FAA ranged from 0.67 (Ser) to 0.92 (Asp) and the bias ranged from -2.0689 µg/mL (Glu) to -0.0487 µg/mL (total FAA).
- The Pearson correlations among the gold standard traits were generally comparable to the Pearson correlations among the respective MIRS-predicted traits.
- Rennet coagulating time was positively associated with MIRS-predicted protein in early lactation (r = 0.19), but was negatively correlated with MIRS-predicted protein in late lactation (r = −0.11), corresponding with the increase in protein concentration across lactation.
- In conclusion, MIRS is useful to routinely and efficiently measure some milk protein fractions and some FAA at a population level.

7.1.3 Chapter 3: Effectiveness of mid-infrared spectroscopy to predict the colour of bovine milk and the relationship between milk colour and traditional milk quality traits

Objective: To evaluate the ability of MIRS to predict milk colour-related traits and to estimate the correlations between these milk colour traits and a selection of traditional milk quality traits.

- Milk colour was measured using a Chroma Meter CR400 with a closed cone set on the L* a*b* system for 601 milk samples from seven research farms.
- Prediction models using MIR were developed to predict each trait separately using partial least squares regression.
- Moderate accuracy of prediction was obtained for the b* index (r_c=0.74 and r_v = 0.72), whereas poor prediction accuracy of prediction was obtained for both the L* index (r_c=0.63 and r_v = 0.55, respectively) and a* index (r_c=0.37 and r_v = 0.30).
- The accuracy of predicting the b* index was greater ($P < 0.05$) when all breeds combined were used in external validation (r_v = 0.72) compared to when just the Holstein-Friesian (r_v = 0.64) or Jersey (r_v =0.66) was used.
- Jersey cows had a greater (P<0.01) mean value for the yellow colour of milk (b* = 10.03) than the Holstein-Friesian cows (b* = 7.48) and their milk also had a greater fat content.
- The linear regression coefficient of the gold standard values on the respective MIRS-predicted values of a*, L*, and b* was 0.81 (0.11), 0.88 (0.05), and 0.96 (0.04), respectively; only the regression coefficient on L* was different from 1.
- The bias ranged from -0.005 (b*) to 0.02 (L*).
- A moderate correlation (0.56) existed between the MIRS-predicted L* and b* indices, both of which were weakly correlated with the a* index.
- The colour traits b* and L* were moderately correlated with MIRS-predicted milk fat, protein, and casein content; the correlation between the gold standard b* and MIRS-predicted fat (0.65) and between the MIRS-predicted b* and MIRS-predicted fat (0.59) were similar.
- In conclusion, MIRS data could be used as a screening tool to efficiently determine the b* colour of milk at a population level, providing a useful tool for the dairy industry and aiding in feeding management and selective breeding.

7.1.4 Chapter 4: Factors associated with protein fractions and FAA predicted using mid-infrared spectroscopy in bovine milk

Objective: To determine the cow and herd level factors associated with detailed protein and FAA composition of bovine milk predicted using MIRS.

- Spectral data used consisted of
    - 94,286 separate morning and evening milk samples from seven research herds.
    - 40,260 milk samples (morning and evening milk samples combined) from 69 commercial herds.
- Mid-infrared spectroscopy prediction models developed in **Chapter 2** were applied to all spectra.
- Factors considered in the linear mixed model included the fixed effects of calendar month of milk test, milking time, parity, stage of lactation, the interaction between parity and stage of lactation, breed proportion of the cow and general heterosis and recombination loss coefficients of the cow; random effects of contemporary group as well as both within and across lactation effects were also fitted.
- When adjusted for crude protein content, total CN and protein fractions (except for α-LA) decreased post-calving to between 36 and 65 DIM across all parties and gradually increased thereafter; the observed decline in total protein and protein fractions in early lactation coincides with the period of negative energy balance observed in early lactation dairy cows.
- A peak in the concentration of all CN fractions was evident in the months of August, September and October. The concentration of Glu was greatest during the months of February, March, April and June when adjusted for milk yield.
- Younger animals produced more total CN, total whey and total β-LG in early and mid-lactation and more Glu and Asp in milk across lactation than their older contemporaries
- Jersey cows produced milk that had a greater concentration of all CN fractions but a lower concentration of total FAA than Holstein cows.
- In conclusion, changes in individual milk protein fractions and FAA across calendar months of the year and across stages of lactation could provide useful input parameters for decision support tools in the management of product portfolios by processors over time.

7.1.5 Chapter 5: Genetic and non-genetic factors associated with milk colour in dairy cows

Objective: To quantify the contribution of cow-level genetic and non-genetic factors to variability in milk colour as described by L*, a* and b* indices predicted using MIRS.

- Spectral data used consisted of 136,807 milk samples from 9,824 cows.
- Mid-infrared spectroscopy prediction models previously developed in **Chapter 2** were applied to all spectra to predict milk lightness (L*), red-green index (a*), and yellow-blue index (b*).
- Factors considered in the linear mixed model included the fixed effects of calendar month of milk test, milking time, parity, stage of lactation, the interaction between parity and stage of lactation, breed proportion of the cow, individual heterosis coefficient and recombination loss among breeds; random effects of contemporary group as well as both within and across lactation effects were also fitted.
- Factors associated with milk colour were breed, stage of lactation, parity, milking-time, somatic cell count, calendar month of the year and season of calving.
- (Co) variance components for $\hat{L}$*, â*, and $\hat{b}$* were estimated using random regressions on the additive genetic and within-lactation permanent environmental effects.
- Heritability estimates varied between 0.15 ± 0.02 (30 DIM) and 0.46 ± 0.02 (210 DIM) for $\hat{L}$*, between 0.09 ± 0.01 (30 DIM) and 0.15 ± 0.02 (305 DIM) for â*, and between 0.18 ± 0.02 (21 DIM) and 0.56 ± 0.03 (305 DIM) for $\hat{b}$*.
- A greater b* predicted value was evident in milk from Jersey cows.
- Milk $\hat{b}$* deteriorated until 31 to 60 DIM, but then improved until the end of lactation.
- Relative to multiparous cows, milk yielded by primiparous cows was, on average, lighter (i.e., greater $\hat{L}$*), more reddish (i.e. greater â*), and less yellow (i.e. lower $\hat{b}$*).
- Strong positive genetic correlations existed between the predicted $\hat{b}$* value and milk fat concentration, ranging from 0.82 ± 0.19 at 5 DIM to 0.96 ± 0.01 at 305 DIM.
- In conclusion, potential exists to breed for different milk colour depending on the respective market demands.

7.1.6 Chapter 6: Genetic parameters of protein fractions and free amino acids predicted using mid-infrared spectroscopy

Objective: to quantify the potential of breeding for improved milk protein and FAA composition, but doing so by exploiting the prediction of these components from routinely available MIRS collected from individual cows at milk testing.

- Spectral data used consisted of
    - 94,286 separate morning and evening milk samples in seven research herds.
    - 40,260 milk samples (morning and evening milk samples combined) in 69 commercial herds.
    - Pedigree information for 33,949 animals was available.
- Gold standard data (715 milk samples) were used to generate equations to predict individual and groups of milk proteins and FAA using the mid-infrared spectrum of milk.
- Genetic, permanent environmental and residual (co) variances for protein fractions and FAA composition were quantified using linear mixed models.
- Heritability estimates for the MIRS-predicted protein ranged from 0.19 for α-LA to 0.46 for β-LG A.
- Heritability estimates for MIRS-predicted FAA ranged from 0.15 for Gly to 0.36 for Asp.
- The estimated genetic standard deviation for each protein fraction trait genetically independent of protein content was less than the respective unadjusted measure; this was also reflected in lower heritability estimates.
- There was little impact on the heritability estimates for FAA when adjusted for differences in the genetic merit of 24 hour milk yield.
- Genetic correlations among the MIRS-predicted protein fractions were weak to strong and ranged from -0.03 (β-LG A and β-LG A) to 0.99 (total CN and α_{S2}-CN).
- Genetic correlations among the MIRS-predicted FAA were also weak to strong and ranged from -0.44 (Asp and Lys) to 0.97 (Glu and total FAA); adjusting the correlations between MIRS-predicted FAA for the genetic merit of 24 hour milk yield did not greatly affect the correlations.
- In conclusion, exploitable genetic variation for protein fractions and FAA exists; these traits can be included in the selection index at no marginal cost, because individual cow (and bulk tank) milk samples are routinely subjected to MIRS analysis.

7.2 Overall conclusions and implications

The overall aim of this thesis was to determine the feasibility of breeding for improved milk quality in dairy cows. In order to breed for a trait such as milk quality, that trait must be: (i) economically or socially important (ii) exhibit genetic variation (i.e. heritable) and (iii) measurable or genetically correlated with a measurable trait. Detailed milk product quality is not considered in the Irish national dairy cow-breeding objective, at present, despite its fundamental importance to potentially add value to the Irish agri-food industry. Inclusion of milk quality traits in national breeding goals is however particularly important for exporting countries, such as Ireland, to consistently achieve a value added high-quality product suitable for international markets.

7.2.1 Genetic variation of milk quality traits

The application of MIRS predictions in breeding programs depends upon the genetic correlation between the predicted and measured values and whether genetic variation in the traits exists. There is a practical utility in the use of MIRS models, if the correlation and the genetic variance of MIRS phenotypes are moderate to strong with reasonable accuracies of prediction by MIRS. The success of breeding programs for increased milk yield in dairy cows is well recognised (Dillon et al., 2006; Norman and Powell, 1999); therefore based on the heritability estimates of colour in **Chapter 5** (predicted a*, L* and b* values were 0.10, 0.29 and 0.35, respectively) and protein fractions and FAA in **Chapter 6** (predicted values of 0.46 for β-LG B and 0.32 for Glu) as well as the coefficient of genetic variation indicates it may be possible to breed for improved colour, protein fractions and FAA in milk.

7.2.2 Phenotyping for milk quality traits

Detailed milk product quality is not considered in the Irish national dairy cow-breeding objective, largely due to lack of routine access to data on milk quality parameters. This is possibly owing to the expense of generating such data using currently available gold standard methods. The use of MIRS as a tool to predict detailed milk measures is attractive since the MIR spectrum is available at no additional cost to routine milk recording and it is faster than the usual gold standard (e.g., Liquid Chromatography-/HPLC-based) techniques. The feasibility of MIRS to predict detailed milk composition traits such as fatty acids (De Marchi et al., 2011; Soyeurt et al., 2011), coagulation traits (De Marchi et al., 2013), as well as animal-level characteristics such as energy balance

(McParland et al., 2011, 2012) and feed efficiency (McParland et al., 2014) has recently been documented. The potential of milk MIRS to predict protein fractions and FAA (**Chapter 2**) as well as milk colour (**Chapter 3**) provides an opportunity to generate large quantities of data for use in genetic evaluations.

7.2.3 Economic importance of milk quality traits

Infant formula production is the fastest growing sector in the world dairy market (FAOSTAT, 2014). Currently, infant formula production contributes €620 million to the value of Irish dairy exports, while cheese contributes €600 million and butter €542 million (Teagasc Publication, 2013). Human and bovine milk have different milk composition; the whey to casein protein ratio varies from 60:40 in human milk to 20:80 in bovine milk. Global cheese production and demand has increased in recent years and this trend is expected to continue (FAOSTAT, 2014). Milk protein composition is important as it affects both yield and characteristics of cheese and plays a vital role in the production of all cheese types (De Marchi et al., 2009a). Higher concentrations of all CN fractions in milk significantly increase cheese yield (Wedholm et al., 2006) and rennet coagulating time is positively correlated with the content and proportions of α_{S1}-CN and α_{S2}-CN in total CN (Bonfatti et al., 2011b). Raw milk colour influences the colour of subsequent dairy products and by-products (Descalzo et al., 2012) thereby influencing the attractiveness of the milk for different markets. A yellow colour of dairy products is considered unfavourable in Middle Eastern dairy markets (Keen and Wilson, 1992), but in Europe, a yellow colour is favourable in high fat dairy products such as butter and full fat cheeses (Hutchings 1994, Casalis et al., 1972).

Milk from certain herds could be selected based on it's protein, FAA or colour profile and perhaps premiums paid for milk profiles that better fit the processors' requirements. Herd-level estimates of milk quality is readily obtained as a by-product from national genetic evaluations and thus the data can be readily available; these herd solutions will be independent of genetic merit of the cows and therefore will more closely reflect the management influence on milk quality. Moreover, being able to monitor the trend in milk quality over time within a herd will provide decision support information to producers on the factors affecting the quality of their milk.

Traditionally, bio-economic models based on the present or the future expected market value or costs of production have been used to define breeding goals for the dairy sector (Veerkamp et al., 2002). Bio-economic models are suitable where the futuristic

profit accruing from a one-unit change in the trait can be accurately measured; this is relatively straightforward for most agro-economic traits (e.g., milk yield, fertility) but can be more difficult for novel traits such as protein fractions and FAA that have no explicit market value (Henchion et al., 2016). Another method to assist in determining the relative emphasis that should be placed on a trait where no market values exist is to undertake a survey of stakeholders to gauge their perceived importance of specific traits. An example of this method is the Delphi technique; the objective of this technique is to (a) identify quality traits that stakeholders think should be included in the national dairy cow breeding goal; (b) understand why stakeholders consider these traits to be important; and (c) direct them towards an agreement on the quality traits that they consider should be included into the national breeding goal. A study was already undertaken using this technique to assess stakeholders' opinion of the importance of detailed milk quality traits within an overall dairy breeding goal for profit; stakeholders included researchers, breeding companies and advisors (Henchion et al., 2016). Results indicated for researchers and processors, milk composition to be the most important attribute, followed closely by protein composition. The majority of farm advisors ranked protein composition in their top three most important attributes. Breeding companies placed a particular emphasis on protein composition. This indicates that across the dairy industry, protein composition is considered important and valuable enough to include in the national breeding objective. Based on the selection theory index, a 16% emphasis should be placed on product quality in the national breeding objective to halt any deterioration in the trait, while based on the Delphi study, a lower emphasis of 4-10% should be placed on product quality; the differences in results indicates the advantage of using more than one method to determine breeding goals (Henchion et al., 2016).

7.2.3 Future research

Several studies have examined the effects of genetic polymorphisms on the content of α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, and β-LG (e.g., Ng-Kwai-Hang et al., 1987; Bobe et al., 1999; Heck et al., 2009) and the technological properties of milk (Ng-Kwai-Hang, 1997; Hill et al., 2002). Selectively breeding for cows with both the β-LG genotype B and the β-κ-CN haplotype A2B will result in milk that is more suitable for cheese production (Heck et al., 2009). The genes coding for milk proteins have been examined in dairy cattle, and a noticeable genetic variation has been identified (Caroli et al., 2009). The major milk proteins of the bovine milk (α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-

LA, and β-LG) are coded by six structural genes (Martin et al., 2002). The four casein genes have been mapped on chromosome 6 in dairy cattle (Hayes and Petit, 1993), whereas the genes encoding α-LA and β-LG have been mapped on chromosomes 5 and 11 (Hayes and Petit, 1993; Hayes et al., 1993).

A genome wide association study is used to identify specific genetic variants associated with a trait; it usually focuses on associations between single nucleotide polymorphisms and traits of importance and variations (Bush et al., 2012). The aims of a genome-wide association studies (GWAS) in dairy cattle breeding is i) to have a better understanding of how genes control production traits ii) to identify markers associated with production traits, these markers can than be incoropated ino genomic evaluations to improve the accuracy of breeding values (Pryce et al., 2010). In recent years, developments in DNA-based marker technology allow the identification of genomic regions (quantitative trait loci, QTL) putatives associated with complex traits such as milk yield and milk composition in dairy cattle. Numerous GWAS for milk production traits on cattle have been completed (Thaller et al., 2003; Weikard et al., 2005; Meredith et al., 2012; Raven et al., 2014). Researchers (e.g., Grisart et al., 2002; Thaller et al., 2003; Pryce et al., 2010) found a causative mutation *(*K232A) on the gene DGAT1 influencing milk production traits including fat and protein on bovine chromosome 14. A GWAS for milk protein composition in dairy cattle was completed by Schopen et al. (2011) and thirty one genomic regions on 20 bovine autosomes were found to be associated with milk protein composition. The region on BTA 6 was associated with all milk proteins and the region on BTA 11 associated with all milk proteins except α-LA.

In comparison to traditional breeding programmes which only used phenotypic and pedigree information for animal evaluation, the inclusion of known quantitive trait loci in genetic evaluations should improve selection accuracy. Producers can selectively breed for cows with desired milk composition at a low cost, creating an opportunity to increase herd profitability (Lopez-Villalobos et al., 2012). A GWAS could be performed on the MIRS predicted dataset used in this thesis to identify genomic regions associated with individual protein composition, FAA and colour. These variants could be incorporated onto the Irish custom genotype panel and used in gentic evaluations to increase accuracy of selection for milk quality traits.

8. References

Abdi, H., M. Lewis-Beck, A. Bryman, and T. Futing. 2003. Partial Least Squares Regrssion (PLS regression). Encyclopedia of Social Sciences Research Methods 25:1-6.

Agabriel, C., A. Cornu, C. Journal, C. Sibra, P. Grolier, and B. Martin. 2007. Tanker milk variability according to farm feeding practices: vitamins A and E, carotenoids, color, and terpenoids. Journal of Dairy Science 90:4884-4896.

Agostoni, C., B. Carratu, C. Boniglia, E. Riva, and E. Sanzini. 2000. Free amino acid content in standard infant formulas: comparison with human milk. Journal of the American College of Nutrition 19(4):434-438.

Albanell, E., C. Gerardo, S. Xavier, R. Maristela, S. Ahmed, and C. Ramón. 2003. Determination of fat, protein, casein, total solids, and somatic cell count in goat's milk by near-infrared reflectance spectroscopy. Journal of the Association of Official Agricultural Chemists International 86:746-752.

Amnarttrakul, R., and A. Thongteeraparp. 2011. New statistics for detection of outliers using the last few principal components. Science Asia 37:355-359.

Andrews, F.J., and R.D. Griffiths. 2002. Glutamine: essential for immune nutrition in the critically ill. British Journal of Nutrition 87:3-8.

Armstrong, M.D., and K.N. Yates. 1963. Free amino acids in milk. Proceedings of the Society for Experimental Biology and Medicine 113:680-683.

Askanazi, J., YA. Carpentier, CB. Michelse, DH. Elwyn, P. Furst, LR. Kantrowitz, FE. Gump, and JM. Kinney. 1980. Muscle and plasma amino acids following injury. Annals of Surgery 192:78-85.

Auldist, M.J., K.A. Johnston, N.J. White, W.P. Fitzsimons, and M.J. Boland. 2004. A comparison of the composition, coagulation characteristics and cheesemaking capacity of milk from Friesian and Jersey dairy cows. Journal of Dairy Research 71:51-57.

Auldist, M.P., S. Coats, G.L. Rogers, and G.H. McDowell. 1995. Changes in composition of milk from healthy and mastitic cows during the lactation cycle. Australian Journal of Experimental Agriculture 35:427-436.

Back, P.J., and N. Lopez-Villalobos. 2007. Breed and heterosis effects for milk protein composition estimated in two stages of lactation in New Zealand dairy cows. Proceedings of the New Zealand Society of Animal Production 67:399-402.

Baker, E.N., and H.M. Baker. 2005. Molecular structure, binding properties and dynamics of lactoferrin. Cellular and Molecular Life Sciences 62:2531-2539.

Bastin, C., H. Soyeurt, and N. Gengler. 2012. Genetic parameters of milk production traits and fatty acid contents in milk for Holstein cows in parity 1-3. Journal of Animal Breeding and Genetics 130:118-127.

Bastin, C., N. Gengler, and H. Soyeurt. 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. Journal of Dairy Science 94:4152-4163.

Berry, D.P., E. Wall, and J.E. Pryce. 2014. Genetics and genomics of reproductive performances in dairy and beef cattle. Animal 8:105-121.

Berry, D.P., J.F. Kearney, K. Twomey, and R.D. Evans. 2013. Genetics of reproductive performance in seasonal calving dairy cattle production systems. Irish Journal of Agriculture and Food Research 52:1-16.

Berry, S.D., S.R. Davis, E.M. Beattie, N.L. Thomas, A.K. Burrett, H.E. Ward, A.M. Stanfield, M. Biswas, A.E. Ankersmit-Udy, P.E. Oxley, J.L. Barnett, J.F. Pearson, Y. van der Does, A.H.K. MacGibbon, R.J. Spelman, K. Lehnert, and R.G. Snell. 2009. Mutation in bovine β-carotene oxygenase 2 affects milk color. Genetics 182:923-926.

Berry, D.P. 2008. Genetics-a tool to improve productivity and profitability. International Journal of Dairy Technology 61:30-35.

Berry, D.P., B. O'Brien, E.J. O'Callaghan, K.O. Sullivan, and W.J. Meaney. 2006. Temporal trends in bulk tank somatic cell count and total bacterial count in Irish dairy herds during the past decade. Journal of Dairy Science 89:4083-4093.

Berry, D.P., R.F. Veerkamp, and P. Dillon. 2006. Phenotypic profiles for body weight, body condition score, energy intake, and energy balance across different parties and concentrate feeding levels. Livestock Science 104:1-12.

Bevilacqua, C., J.C. Helbling, G. Miranda, and P. Martin. 2006. Translational efficiency of casein transcripts in the mammary tissue of lactating ruminants. Reproduction Nutrition Development 46:567-578.

Bjugn, R., and B. Casati. 2012. Stakeholder analysis: a useful tool for biobank planning. Biopreservation and Biobanking 10:239-244.

Bobe, G., D.C. Beitz, A.E. Freeman, and G.L. Lindberg. 1999. Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. Journal of Dairy Science 82:2797-2804.

Bonfatti, V., G. Di Martino, and P. Carnier. 2011a. Effectiveness of mid-infrared spectroscopy for the prediction of detailed protein composition and contents of protein genetic variants of individual milk of Simmental cows. Journal of Dairy Science 94:5776-5785.

Bonfatti V, A. Cecchinato, L. Gallo, A. Blasco, and P Carnier. 2011b. Genetic analysis of detailed milk protein composition and coagulation properties in Simmential cattle. Journal of Dairy Science 94, 5183–5193.

Brereton, R.G. 2015. The Mahalanobis distance and its relationship to principal components scores. Journal of Chemometrics 29:143-145.

Bramanti, E., C. Sortino, and G. Raspi. 2002. New chromatographic method for separation and determination of denatured αS1-, αS2-, β- and κ-caseins by hydrophobic interaction chromatography Journal of Chromatography 958:157-166.

Bryant, J.R., N. Lopez-Villalobos, J.E. Pryce, C.W. Holmes, D.L. Johnson, and D.J. Garrick. 2007. Short Communication: Effect of environment on the expression of breed and heterosis effects for production traits. Journal of Dairy Science 90:1548-1553.

Buckley, F., N. Lopez-Villalobos, and B.J. Heins. 2014. Crossbreeding: Implications for dairy cow fertility and survival. Animal 8:122-133.

Bush, W.S., and J.H. Moore. 2012. Chapter 11: Genome-wide association studies. PLOS Computational Biology 8:12.

Calderon, F.B., P. Chaveau-Duriot, B. Pradel, B. Martin, M. Graulet, P. Doreau, and P. Noziere. 2007. Variations in carotenoids, vitamins A and E, and colour in cow's plasma and milk following a shift from hay diet to diets containing increasing levels of cartenoids and vitamin E. Journal of Dairy Science 90:5651-5664.

Caroli, A.M, S. Chessa, G.J. Erhardt. 2009. Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. Journal of Dairy Science 92:5335-5352.

Cassady, J.P., L.D. Young, and K.A. Leymaster. 2002. Heterosis and recombination associations on pig reproductive traits. Journal of Animal Science 80:2303-2315.

Casalis, J., Y. Chardon, F.M. Luquet, P. Mainguy, and M. Yver. 1972. Sur la relation entre la teneur en beta-carotene et la couleur des beurres francais. Lait 52:28-42.

Cassandro, M.A.M. Comin, R. Ojala, M. Dal Zotto, M. De Marchi, L. Gallo, P. Carnier, and G. Bittante. 2008. Genetic parameters of milk coagulation properties and their relationships with milk yield and quality traits in Italian Holstein cows. Journal of Dairy Science 91:371-375.

Cecillon, L., B.G. Barthes, C. Gomez, D. Ertlen, V. Genot, M. Hedde, A. Steven, and J.J. Brun. 2009. Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). Journal of Soil Science 60:770-784.

Cerbulis, J., and H.M. Farrell. 1975. Composition of milks of dairy cattle. I. Protein, Lactose, and Fat Contents and Distribution of Protein Fraction. Journal of Dairy Science 58:817-827.

CIELAB. 1976. CIELAB Colour System. Commission International de l'Eclairage, Paris, France.

Chang, W.C., D.A. Laird, M.J. Mausbach, and C.R. Hurburgh. 2001. Near-Infrared reflectance spectroscopy–Principal components regression analyses of soil properties. Soil Science society of America Journal 65:480-490.

Chuang, C.K., S.P. Lin, H.C. Lee, T.J. Wang, Y.S. Shih, K.Y. Huang, and C.Y. Yeung. 2005. Free amino acids in full-term and pre-term human milk and infant formula. Journal of Pediatric and Gastroenterology Nutrition 40:496-500.

Coffey, E.L., B. Horan, R.D. Evans, and D.P. Berry. 2016. Milk production and fertility performance of Holstein, Friesian, and Jersey purebred cows and their respective crosses in seasonal-calving commercial farms. Journal of Dairy Science 99:5681-5689.

Cozzolino. D., M.J. Kwiatkwoski, R.G. Dambergs, W.U. Cynkar, L.J. Janik, G. Skouroumounis, and M. Gishen. 2008. Analysis of elements in wine using near infrared spectroscopy and partial least squares regression. Talanta 74:711-716.

Cozzolino, D., and A. Moron. 2004. Exploring the use of near infrared reflectance spectroscopy (NIRS) to predict trace minerals in legumes. Animal Feed Science and Technology 111:161-173.

Cozzolino, D., A. Fassio, and A. Gimenez. 2000. The use of near-infrared reflectance spectroscopy (NIRS) to predict the composition of whole maize plants. Journal of the Science of Food and Agriculture 81:142-146.

Curi, R., P. Newsholme, J. Procopio , C. Lagranha, R. Gorjão, and T.C. Pithon-Curi. 2007. Glutamine, gene expression, and cell function. Frontiers in Bioscience 12:344-357.

Dal Zotto, R., M. De Marchi, A. Cecchinato, M. Penasa, M. Cassandro, P. Carnier, L. Gallo, and G. Bittante. 2008. Reproducibility and repeatability of measures of milk coagulation properties and predictive ability of mid-infrared reflectance spectroscopy. Journal of Dairy Science 91:4103-4112.

Davis, T.A., H.V. Nguyen, R. Garcia-Bravo, M.L. Fiorotto, E.M. Jackson, and P.J. Reids. 1994. Amino acid composition of the milk of some mammalian species changes with stage of lactation. British Journal of Nutrition 72:845-853.

Davies, D.T, and J.C.D. White. 1966. The stability of milk protein to heat: I. Subjective measurement of heat stability of milk. Journal of Dairy Research 33:67-81.

Dechow, C.D., G.W. Rogers, J.B. Cooper, M.I. Phelps, and A.L. Mosholder. 2007. Milk, fat, protein, somatic cell score, and days open among Holstein, Brown Swiss, and their Crosses. Journal of Dairy Science 90:3542-3549.

Dehareng, F., C. Delfosse, E. Froidmont, H. Soyeurt, C. Martin, N. Gengler, A. Vanlierde, and P. Dardenne. 2012. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. Animal 6:1694-1701.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as a phenotyping tool for milk traits. Journal of Dairy Science. 97:1171-1186.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2013. Prediction of coagulating and non-coagulating milk samples using mid-infrared spectroscopy. Journal of Dairy Science 96:4707-4715.

De Marchi, M., M. Penasa, A. Cecchinato, M. Mele, P. Secchiari, and G. Bittante. 2011. Effectiveness of mid-infrared spectroscopy to predict fatty acid composition of Brown Swiss bovine milk Animal 5:1653-1658.

De Marchi, M., M. Penasa, M. Battagin, E. Zanetti, C. Pulici, and M. Cassandro. 2011. Feasibility of the direct application of near-infrared reflectance spectroscopy on intact chicken breasts to predict meat colour and physical traits. Poultry Science 90:1594-1599.

De Marchi, M., V. Bonfatti, A. Cecchinato, G. Di Martino, and P. Carnier. 2010. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. Italian Journal of Animal Science 8:399-401.

De Marchi, M., C.C. Fagan, C.P. Donnell, A. Cecchinato, R. Dal Zotto, M. Cassandro, M. Penasa, and G. Bittante. 2009a. Prediction of coagulation properties, titratable acidity, and pH of bovine milk using mid-infrared spectroscopy. Journal of Dairy Science 92:423-432.

De Marchi, M., V. Bonfatti, A. Cecchinato, G. Di Martino, and P. Carnier. 2009b. Prediction of protein composition of individual cow milk using mid-infrared spectroscopy. Italian Journal of Animal 8:399-401.

Descalzo, A.M., E. Comerón, E. Salado, G. Bretschneider, G. Gagliostro, G. Grigioni, L. Antonacci, L. Costabel, L. Negri, L. Rossetti, and M.A. Taverna. 2012. Differential characteristics of milk produced in grazing systems and their impact on dairy products. INTECH Open Access Publisher.

Department of Agriculture Fisheries and Food (DAFF). Food Harvest 2020 A Vision for Irish Agri-Food and Fisheries. Accessed June 10, 2016. http://www.agriculture.gov.ie/media/migration/agrifoodindustry/foodharvest2020/foodharvest2020/2020strategy/2020Foodharvestl90710.pdf.

De Wit, J.N. 1998. Nutritional and functional characteristics of whey proteins in food products. Journal of Dairy Science 81:579-607.

Dillon, P., D.P. Berry, R.D. Evans, F. Buckley, and B. Horan. 2006. Consequences of genetic selection for increased milk production in European seasonal pasture based systems of milk production. Livestock Science 99:141-158.

Dousseau, F., M . Therrien, and M. Pézolet. 1989. On the spectral subtraction of water from the FT-IR spectra of aqueous solutions of proteins. Applied Spectroscopy 43:538-542.

Dziki, W., and J. Doddi. 2008. Pharmaceutical applications of mid-infrared spectroscopy in GMP compliance. Compliance Methods 12:48-55.

Elofsson, U.M., M.A. Paulsson, P. Sellers, and T. Arnebrant. 1996. Adsorption during heat treatment related to the thermal unfolding/aggregation of b-lactoglobulins A and B. Journal of Colloid and Interface Science 183:408-415.

Erdman, R.A., and M. Varner. 1995. Fixed yield responses to increased milking frequency. Journal of Dairy Science 78:1199-1203.

Espada, E., and H. Vijverberg. 2002. Milk colour analysis as a tool for the detection of abnormal milk. North American Conference on robotic milking, Toronto, Canada 28-38.

Etzion, Y., R. Linker, U. Cogan, and I. Shmulevich. 2004. Determination of protein concentration in raw milk by mid-infrared Fourier Transform infrared/attenuated total reflectance spectroscopy. Journal of Dairy Science 87:2779-2788.

Everett, R.W., and L.H. Wadell. 1970. Sources of variation affecting the difference between morning and evening daily milk production. Journal of Dairy Science 53:1424-1429.

FAOSTAT. 2014. Metadata. Accessed September 11, 2014. http://faostat3.fao.org/faostat-gateway/go/to/download/Q/*/E.

Farrell, H.M., R. Jimenez-Flores, G.T. Bleck, E.M. Brown, J.E. Butler, L.K. Creamer, C.L. Hicks, C.M. Hollar, K.F. Ng-Kwai-Hang, and H.E. Swalsgood. 2004. Nomenclature of the proteins of cows' Milk-Sixth Revision. Journal of Dairy Science 87:1641-1674.

Farnaud, S., and R.W. Evans. 2003. Lactoferrin–a multifunctional protein with antimicrobial properties. Molecular Immunology 40:395-405.

Ferrand, M., B. Huquet, S. Barbey, F. Barillet, F. Faucon, H. Larroque, O. Leray, J. M. Trommenschlager, and M. Brochard. 2011. Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: interest of applying a variable

selection by genetic algorithms before a PLS regression. Chemometrics and Intelligent Labortory Systems 106:183-189.

Ferlay, A., B. Martin, S. Lerch, M. Gobert, P. Pradel, and Y. Chilliard. 2010. Effects of supplementation of maize silage diets with extruded linseed, vitamin E and plant extracts rich in polyphenols, and morning v. evening milking on milk fatty acid profiles in Holstein and Montbeliarde cows. Animal 4:627-640.

Fletcher, D.L. 1999. Broiler breast meat color variation, pH, and texture. Poultry Science 78:1323-1327.

Fox, P.F., and P.L. McSweeney. 1998. Dairy chemistry and biochemistry. Springer 105-111.

Ghadimi, H., and P. Pecora. 1963. Free amino acids of different kinds of milk. American Journal of Clinicial Nutrition 13:75-81.

Gilbert, G.R., G.L. Hargrove, and M. Kroger. 1973. Diurnal variations in milk yield, fat yield, milk fat percentage, and milk protein percentage of Holstein-Friesian cows. Journal of Dairy Science 56:409-410.

Gilmour, A.R., B.R. Cullis, S.J. Welham, and R. Thompson. 2011. ASREML Reference Manual. VSN International Ltd., Hemel Hempstead, UK.

Gilmour, A.R., R.D. Anderson, and A.L. Rae. 2009. The analysis of binomial data by a generalized linear mixed model. Biometrik Oxford Journals.

Gleeson, M. 2008. Dosing and efficacy of glutamine supplementation in human exercise and sport training. Journal of Nutrition 138:2045S-2049S.

Graml, R., and F. Pirchner. 2003. Effects of milk protein loci on content of their proteins. Archiv fur Tierzucht 46:331-340.

Grisart, B., W. Coppitiers, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Research 12:222-231.

Gross, J.J., E.C. Evelyne, C. Kesslera, and R.M. Bruckmaier. 2014. Colour measurement of colostrum for estimation of colostral IgG and colostrum composition in dairy cows. Journal of Dairy Research 81:400-444.

Gutiérrez-Adán, A., E.A. Maga, H. Meade, C.F. Shoemaker, J.F. Medrano, G.B. Anderson, and J.D. Murray. 1996. Alterations of the physical characteristics of milk from transgenic mice producing bovine κ-Casein. Journal of Dairy Science 79:791-799.

Ha, E., and M.B. Zemel. 2003. Functional properties of whey, whey components, and essential amino acids: mechanisms underlying health benefits for active people (review). The Journal of Nutritional Biochemistry 14:251-258.

Habuka, P., Y. Murakamig, M. Noma, T. Kudotl, and K. Horikoshill. 1989. Amino acid sequence of *Mirabilis* antiviral protein, total synthesis of its gene and expression in *Esclierichiu coli.* The Journal of Biological Chemistry 264:25.

Hall, R.L. 1970. Pasture development in spear grass region at Westwood in the fitzroy basin. Tropical Grasslands 4:77-84.

Hayes, H.C, and E.J. Petit. 1993a. Mapping of the beta-lactoglobulin gene and of an immunoglobulin-M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. Mammalian Genome 4:207-210.

Hayes, H.C, E.J. Petit, C. Bouniol, and P. Popescu. 1993. Localisation of the alpha-S2-casein gene (CASAS2) to the homologous cattle, sheep and goat chromosomes 4 by in situ hybridization. Cytogenetics and Cell Genetics 64:282-5.

Haug, A., A.T. HØstmark, and O.M. Harstad. 2007. Bovine milk in human nutrition-a review. Lipids in Health and Disease 6:25.

Heck, J.M.L., H.J.F. Van Valenberg, J. Dijkstra, and A.C.M Van Hooijdonk. 2009. Seasonal variation in the Dutch bovine raw milk composition. Journal of Dairy Science 92:4745-4755.

Helland, I.S. 1990. Pls regression and statistical models. Scandivian Journal of Statistics 17:97-114.

Henchion, M., M. McCarthy, V.C. Resconi, D.P. Berry, and S. McParland. 2016. Stakeholder involvement in establishing a milk quality sub-index in dairy cow breeding goals: a Delphi approach. Animal 10:878-891.

Hettinga, K.A., H.J.F. Van Valenberg, T.J.G.M. Lam, and A.C.M. Van Hooijdonk. 2008. Detection of mastitis pathogens by analysis of volatile bacterial metabolites. Journal of Dairy Science 91:3834-3839.

Hill, J.P., M.J. Boland, P. Harris, and G. Paterson. 2000. Impact of genetic polymorphism on milk powder manufacture and processing. In: Agnew RE, Agnew KW, Fearon AM, editors. Milk Composition. British Society of Animal Science 87-92.

Hollar, C.M., A.J.R. Law, D.G. Dalgleish, and R.J. Brown. 1991. Separation of major casein fractions using cationexchange fast protein liquid chromatography. Journal of Dairy Science 74:2403-2409.

Huang, W., F. Penagaricano, K.R. Ahmad, J.A. Lucey, KA. Weigel, and H. Khatib. 2012. Association between milk protein gene variants and protein composition traits in dairy cattle. Journal of Dairy Science 95:440-449.

Hutchings, J.B. 1994. Food Colour and Appearance. Blackie Academic, London, UK.

Irish Co-Operative Organisation society. Last accessed June 25th, 2016 http://www.icos.ie/members/dairy

Hutton, K., R.C. Seeley, and D.G. Armstrong. 1969. The variation throughout a year in the fatty acid composition of milk fat from 2 dairy herds. Journal of Dairy Research 36:103-113.

ICOS (The Irish Co-operative Organisation Society). 2016. Last Accessed July 10th, 2016. http://www.icos.ie/about/purpose-vision

Ikonen, T., S. Morry, A.M. Tyriseva, O. Routtine, and M. Ojala. 2004. Genetic and phenotypic correlations between milk coagulation properties, milk production traits, somatic cell count, casein content and pH of milk. Journal of Dairy Science 87:458-467.

Ikonen, T., K. Ahlfors, R. Kempe, M. Ojala, and O. Ruottinen. 1999. Genetic Parameters for the Milk Coagulation Properties and Prevalence of Noncoagulating Milk in Finnish Dairy Cows. Journal of Dairy Science 82:205-214.

Indyk, H. 1987. The rapid determination of carotenoids in bovine milk using HPLC. Journal of Micronutritient Analysis 3:169-183.

Jabed A, S. Wagner, J. McCracken, DN. Wells, and G. Laible. 2012. Targeted microRNA expression in dairy cattle directs production of b-lactoglobulin-free, high-casein milk. Proceedings of the National Academy of Sciences 10:16811-16816.

Jadhav, B.S., M.B. Kulkarni, and K.D. Chavan. 2008. Effect of lactation order and stage of lactation on physical properties of milk. Journal of Dairying, Foods and Home Sciences 27:168-174.

Jensen, R.G., B. Blanc, and S. Patton. 1995. Particulate constituents in human and bovine milks. Academic Press: San Diego, CA, USA 50-62.

Jimidar, M, T.P. Hamoir, A. Foriers, and D.L. Massart. 1993. Comparison of capillary zone electrophoresis with high-performance liquid chromatography for the determination of additives in foodstuffs. Journal of Chromatography A 636:179-186.

Jolliffe, I. 2002. Principal Component Analysis. Springer 7:50.

Joudu, I., M. Henno, T. Kaart, T. Pussa, and O. Kart. 2008. The association of milk protein contents on the rennet coagulation properties of milk from individual dairy cows. International Dairy Journal 18:964-967.

Kawasaki, M., S. Kawamura, M. Tsukahara, S. Morita, M. Komiya, and M. Natsuga. 2008. Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot. Computers and Electronics in Agriculture 63:22-27.

Keen, A.R., and R.D. Wilson. 1992. Effect of breed on colour and flavour. Summary of Proceedings of the Milk Fat Flavour Forum 50-54.

Kelsey, J.A., B.A. Corl, R.J. Collier, and D.E. Bauman. 2003. The Effect of Breed, Parity, and Stage of Lactation on Conjugated Linoleic Acid (CLA) in Milk Fat from Dairy Cows. Journal of Dairy Science 86:2588-2597.

Kim, H.Y., and R. Jimenez-Florez. 1994. Comparison of milk proteins using preparative isoelectric focusing followed by polyacrylamide gel electrophoresis. Journal of Dairy Science 77:2177-2190.

Klei L.R., J.M. Lynch, D.M. Barbano, P.A. Oltenacu, A.J. Lednor, and D.K. Bandler. 1997. Influence of milking three times a day on milk quality. Journal of Dairy Science 80:427-436.

Koletzko, B., S. Baker, G. Cleghorn, U.F. Neto, S. Gopalan, O. Hernell, S.Q. Hock, P. Jirapinyo, B. Lonnerdal, P. Pencharz, H. Pzyrembel, J. Ramirez-Mayans, R. Shamir, D. Turck, Y. Yamashiro, and D. Zong-Yi. 2005. Global standard for the composition of infant formula: Recommendations of an ESPGHAN coordinated international expert group. Journal of Pediatric Gastroenterology and Nutrition 41:584-599.

Kroeker, E.M., K.F. Ng-Kwai-Hang, J.F. Hayes, and J.E. Moxley. 1985. Associations of Environmental factors and milk protein polymorphism on composition of casein fraction in bovine milk. Journal of Dairy Science 68:1752-1757.

Kunz, C and B. Lönnerdal. 1992. Re-evaluation of the whey protein/casein ratio of human milk. Acta Paediatrica. 8:107-112.

Larsen, M.K., U., Kidmose, T., Kristensen, P., Beaumont, and G. Mortensen. 2013. Chemical composition and sensory quality of bovine milk as affected by type of forage and proportion of concentrate in the feed ration. Journal of the Science of Food and Agriculture 93:93-99.

Laugesen, M., and R. Elliott. 2003. Ischaemic heart disease, type 1 diabetes, and cow milk A1 beta-casein. New Zealand Medical Journal 116:1-19.

Lien, E.L. 2003. Infant formulas with increased concentrations of α-lactalbumin. The American Journal of Clinical Nutrition 77:1555-1558.

Lindmark-Mansson, H., R. Fonde, and H.E. Petterson. 2003. Composition of Swedish dairy milk. International Dairy Journal 13:409-425.

Lindström, U.B., V. Antila, and J. Syväjärvi. 1984. A note on some genetic and non-genetic factors affecting clotting time of Ayrshire milk. Acta Agriculturae Scandinavica 34:349-355.

Liu, Y., B.G. Lyon, W.R. Windham, C.E. Realini, T.D. Pringle, and S. Duckett. 2003. Prediction of color, texture, and sensory characteristics of beef steaks by visible and near infrared reflectance spectroscopy. A feasibility study. Meat Science 65:1107-1115.

Lönnerdal, B. 2003. Nutritional and physiological significance of human milk proteins. American Journal of Clinicial Nutrition 77:1537-1543.

Lopez-Villalobos, N. 2012. Analysing the genetic basis of milk production traits. CAB Reviews 5167:1-18.

Lopez-Villalobos, N., S.R. Davis, E.M. Beattie, J. Melis, S. Berry, S.Holroyd, R.J. Spelman, and R.G. Snell. 2009. Breed effects for lactoferrin concentration determined by Fourier transform infrared spectroscopy. Proceedings of New Zealnd Society of Animal Production 69:1-5.

Lucas, A., D., Andueza, E. Rock, and B. Martin. 2008. Prediction of dry matter, fat, pH, vitamins, minerals, carotenoids, total antioxidant capacity, and colour in fresh and freeze-dried cheeses by visible-near-infrared reflectance spectroscopy. Journal of Agriculture and Food Chemistry 56:6801–6808.

Luginbuhl, W. 2002. Evaluation of designed calibration samples for casein calibration in Fourier Transform infrared analysis of milk. LWT-Food Science and Technolology 35:554-558.

Lukin, J.A., and C. Ho. 2004. The structure-function relationship of hemoglobin in solution at atomic resolution. Chemical Reviews 104:1219-1230.

MacGibbon, A.K.H., and M.W. Taylor. 2006. Composition and structure of bovine milk lipids. Advanced Dairy Chemistry 2:1-42.

Martens, H, and T. Naes. 1989. Multivariate Calibration. London:Wiley.

Martin P, M. Szymanowska, L. Zwierzchowski, C. Leroux. 2002. Theimpact of genetic polymorphisms on the protein composition of ruminant milks. Reproduction Nutrition Development 42:433-459.

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.

Meredith, B.K., F.J. Kearney, E.K. Finlay, D.G. Bradley, A.G. Fahey, D.P. Berry, and D.J. Lynn. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genetics 13:21.

McLachlan, CN. 2001. Beta-casein A1, ischaemic heart disease mortality, and other illnesses. Medical Hypotheses 56:262-272.

McLean, DM., ERB. Graham, and R.W. Ponzoni. 1984. Effects of milk protein genetic variants on milk yield and composition. Journal of Dairy Research 51:531-46.

McParland, S. and D.P. Berry. 2016. The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows. Journal of Dairy Science 99:4056-4070.

McParland, S., E. Lewis, E. Kennedy, S.G. Moore, B. McCarthy, M. O'Donovan, S.T. Butler, J.E. Pryce, and D.P. Berry. 2014. Mid-infrared spectroscopy of milk as a predictor of energy intake and efficiency in lactating dairy cows. Journal of Dairy Science 97:5863-5871.

McParland, S., G. Banos, B. McCarthy, E. Lewis, M.P. Coffey, B. O'Neill, M. O'Donovan, E. Wall, and D.P. Berry. 2012. Validation of mid-infrared spectrometry in milk for predicting body energy status in Holstein-Friesian cows. Journal of Dairy Science 95:7225-7235.

McParland, S., G. Banos, E. Wall, M.P. Coffey, H. Soyeurt., and D.P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cow Journal of Dairy Science 94:3651-3661.

Mehrotra, R., and H.W. Siesler. 2003. Application of Mid/Near Infrared Spectroscopy in Sugar Industry. Applied Spectroscopy Reviews 3:307-354.

Mero, A. 1999. Leucine Supplementation and Intensive Training. Sports Medicine 27:347-358.

Miglior, F., B.L. Muir, and B.J. Van Doormaal. 2005. Selection Indices in Holstein Cattle of Various Countries. Journal of Dairy Science 88:1255-1263.

MilkoScan™ FT+ Basic Calibrations Application Note 5373 / Rev. 2 Part No. 6007 0238

Miller, MJS., SA. Witherly, and DA. Clark. 1990. Casein: a milk protein with diverse biologic consequences. Experimental Biology and Medicine 195:143-159.

Mitsumoto, M., S. Maeda, T Mitsuhashi, and S. Ozawa. 1991. Near-infrared spectroscopy determination of physical and chemical characteristics in beef cuts. Journal of Food Science 56:1493-1496.

Miralles, B., J. Leaver, M. Ramos, and L. Amigo. 2003. Mass mapping analysis as a tool for the identification of genetic variants of bovine β-casein. Journal of Chromatography A 1007:47-53.

Mohan, L.A., C. Karunakaran, D.S. Jayas, and N.D.G. White. 2005. Classification of bulk cereals using visible and NIR reflectance characteristics. Canadian Biosystems Engineering 47:7-14.

Mollé, D., and J. Léonil. 2005. Quantitative determination of bovine κ-casein macropeptide in dairy products by Liquid chromatography/Electrospray coupled

to mass spectrometry (LC-ESI/MS) and Liquid chromatography/Electrospray coupled to tamdem mass spectrometry (LC-ESI/MS/MS). International Dairy Journal 15:419-428.

Morales, M.S., D.L. Palmquist, and W.P. Weiss. 2000. Milk fat composition of Holstein and Jersey cows with control or depleted copper status and fed whole soybeans or tallow. Journal of Dairy Science 83:2112-2119.

Morris, C.A., T.W. Knight, S.A.N. Newman, S.M. Hickey, A.F. Death, K.T. O'Neill, and M. Ridland. 2002. Genetic studies of carotenoid concentration in the plasma and milk of New Zealand dairy cattle. New Zealand Journal of Agricultural Research 45:27-33.

Mounier, J., M.C. Rea, P.M. O'Connor, G.F. Fitzgerald, and T.M. Cogan. 2007. Growth characteristics of brevibacterium, corynebacterium, microbacterium, and staphylococcus spp. isolated from surface-ripened cheese. Applied Environmental Microbiology 73:7732-7739.

Mounsey, J.S., and B.T. O'Kennedy. 2009. Stability of β-lactoglobulin/micellar casein mixtures on heating in simulated milk ultrafiltrate at pH 6.0. International Journal of Dairy Technology 62:493-499.

Nakayama, J., J.C. Rice, B.D. Strahl, C.D. Allis, and S.L.S Grewal. 2001. Role of histone NH3 Lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292:110-113.

Nelson, V. 1948. The spectrophotometric determination of the colour of milk. Journal of Dairy Science 31:409-413.

Ng-Kwai-Hang, K.F. 2002. Heterogeneity, fractionation and isolation. Encyclopaedia of Dairy Sciences. London: Academic Press 1881-1894.

Ng-Kwai-Hang KF. 1997. A review of the relationship between milk protein polymorphism and milk composition/milk production. In: Federation ID, editor. Milk Protein Polymorphism. International Dairy Federation, Brussels, Belgium 22-37.

Ng-Kwai-Hang, K.F., and S. Kim. 1996. Different amounts of beta-lactoglobulin A and B in milk from heterozygous AB cows. International Dairy Journal 6:689-695.

Ng-Kwai-Hang, K.F., J.F. Hayes, J.E. Moxley, and H.G. Monardes. 1987. Variation in milk protein concentrations associated with genetic polymorphism and environmental factors. Journal of Dairy Science 70:563-570.

Ng-Kwai-Hang, K.F., and E.M. Kroeker. 1984. Rapid separation and quantification of major caseins and whey proteins of bovine milk by polyacrylamide gel electrophoresis. Journal of Dairy Science 67:3052-3056.

Norman, H.D., and R.L. Powell. 1999. Dairy cows of high genetic merit for yields of milk, fat, and protein-a review. Journal of Animal Science 12:1316-1323.

Noziere, P., B. Graulet, A. Lucas, B. Martin, P. Grolier, and M. Doreau. 2006a. Carotenoids for ruminants: From forages to dairy products. Animal Feed Science Technology 131:418-450.

Noziere, P., P. Grolier, D. Durand, A. Ferlay, P. Pradel, and B. Martin. 2006b. Variations in carotenoids, fat-soluble micronutrients, and colour in cows'

plasma and milk following changes in forage and feeding level. Journal of Dairy Science 89:2634-2648.
Oloffs, K.V., H. Schulte-Coerne, K. Pabst, and H.O. Gravert. 1992. Die Bedeutung der Proteinvarianten für genetische Unterschiede in der Käsereitauglichkeit der Milch. Züchtungskunde 64:20-26.
Oltenacu, P.A., and D.M, Broom. 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. Animal welfare 19:39-49.
Ordolff, D. 2006. Aussagekraft von standard-milchinhaltsstoffen zur bewertung der eutergesundheit. Landtechnik 61:48-49.
Osborne, B.G., R.J. Henry, and M.D. Southan. 2007. Assessment of commercial milling potential of hard wheat by measurement of the rheological properties of whole grain. Journal of Cereal Science 45:122-127.
Ostersen, S., J. Foldager, and J.E. Hermansen. 1997. Effects of stage of lactation, milk protein genotype and body condition at calving on protein composition and renneting properties of bovine milk. Journal of Dairy Research 64:207-219.
Ouweltjes, W. 1998. The relationship between milk yield and milking interval in dairy cows. Livestock Production Science 56:193-201.
Owens, C.J.W., IL. Labuschagne, and MJ. Lombard. 2013. The basics of prescribing infant formulas. South Africian Family Practice 55:123-13.
Parmar, B.L., R.E. Freeman, J.S. Harrison, A.C. Wicks, S. De Colle, and L. Purnell. 2010. Stakeholder theory: the state of the art. The Academy of Management Annals 4:403-445.
Pellizon Birelli, M. and G. Fazio. 2006. La spettroscopia di assorbimento infrarosso, una tecnica analitica per la caratterizzazione chimico fisica dei materiali, Enco - Engineering Concrete, Ponzano Veneto.
Phillips, L.G., M.L. McGiff, D.M. Barbano, and H.T. Lawless. 1995. The Influence of Fat of the Sensory Properties, Viscosity, and Colour of Lowfat Milk. Journal of Dairy Science 78:128-1266.
Ploumi, K., S. Belibasaki, and G. Triantaphyllidis. 1998. Some factors affecting daily milk yield and composition in a flock of Chios ewes. Small Ruminant Research 28:89-92.
Pryce, J.E., S. Bolormaa, A.J. Chamberlain, P.J. Bowman, K. Savin, M.E. Goddard, and B.J. Hayes. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. Journal of Dairy Science 93:3331-3345.
Quist, M.A., S.J. LeBlanc, K.J. Hand, D. Lazenby, F. Miglior, and D.F. Kelton. 2008. Milking-to-milking variability for milk yield, fat and protein percentage, and somatic cell count. Journal of Dairy Science 91:3412–3423.
Raty, J.A., and K.E. Peiponen. 1999. Reflectance study of milk in the UV-visible range. Applied Spectroscopy 53:1123-1127.
Raven, L.A., B.G. Cocks, and B.J. Hayes. 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15:2-14.

Recio, I., L. Amigo, and R. López-Fandiño. 1997. Assessment of the quality of dairy products by capillary electrophoresis of milk proteins. Journal of Chromatography B 697:231-242.

Reeves, JB., and J.S. Van Kessel. 2000. Near-infrared spectroscopic determination of carbon, total nitrogen, and ammonium-N in dairy manures. Journal of Dairy Science 83:1829-1836.

Reid, M., M. O'Donovan, C.T. Elliot, J.S. Bailey, C.J. Watson, S.T.J. Lalor, B. Corrigan, M.A. Fenlon, and E. lewis. 2015. The effect of dietary crude protein and phosphorus on grass-fed dairy cow production, nutrient status and milk heat stability. Journal of Dairy Science 98:1-15.

Rendel, JM and A. Robertson. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. Journal of Genetics 50:1-8.

Renner, E., and U. Kosmack. 1975. Genetische Aspekte zum Eiweisgehalt und zu den Eiweisfraktionen der Milch. II. Eiweisfraktionen. Zuchtungskunde 47:441-457.

Roucher, V.F., E. Desnots, C. Nael, A.M. Agnoux, M.C. Alexandre-Gouabeau, D. Darmaun, and C.Y Boquien. 2013. Use of UPLC-ESI-MS/MS to quantitate free amino acid concentrations in micro-samples of mammalian milk. SpringerPlus 2:622.

Rutten, M.J.M., H. Bovenhuis, J.M.L. Heck, and J.A.M. van Arendonk. 2011. Predicting bovine milk protein composition based on Fourier transform infrared spectra. Journal of Dairy Science 94:5683-5690.

Rutten, M.J.M., H. Bovenhuis, K.A. Hettinga, H.J.F. Van Vanlenberg, and J.A.M. Van Arendonk. 2009. Predicting bovine milk fat composition using infrared spectroscopy based on milk samples collected in winter and summer. Journal of Dairy Science 92:6202-6209.

Sarwar, G., H.G. Botting, T.A. Davis, P. Darling, and P.B. Pencharz. 1998. Free amino acids in milks of human subjects, other primates and non-primates. British Journal of Nutrition 79:129-131.

Schopen, G.C.B., J.M.L. Heck, H. Bovenhuis, M.H.P.W. Visker, H.J.F. van Valenberg, and J.A.M. van Arendonk. 2009. Genetic parameters for major milk proteins in Dutch Holstein-Friesians. Journal of Dairy Science 92:1182-1191.

Schwarz, E.L., W.L. Roberts, and M. Pasquali. 2005. Analysis of plasma amino acids by HPLC with photodiode array and fluorescence detection. Clinica Chimica Acta 354:83-90.

Searle, S.R. 1961. Phenotypic, genetic and environmental correlations. Biometrics 17:474-480.

Shalloo, L., and U. Geary. 2011. Milk Payment. Accessed May 11th, 2016. Available:http://www.teagasc.ie/dairy/financial/docs/milkpayment.pdf.

Siesler, H.W., W. Ozaki, S. Kawata, and H.M. Heise. 2008. Near Infrared Spectroscopy: Principles, instruments, applications. Wiley-VCH Verlag, GmbH D-69469, Weinheim, Germany.

Singh, H. 2004. Heat stability of milk. International Journal of Dairy Technology 57:11-119.

Sklan, D., M. Kaim, U. Moallem, and Y. Folman. 1994. Effect of dietary calcium soaps on milk yield, body weight, reproductive hormones, and fertility in first parity and older cows. Journal of Dairy Science 77:1652-1660.

Solah, V.A., V. Staines, S. Honda, and H.A. Limley. 2007. Measurement of milk colour and composition: Effect of dietary intervention on Western Australian Holstein - Friesian cow's milk quality. Journal of Food Science 72:560-566.

Sørensen, M.K., E. Norberg, J. Pedersen, and L. G. Christensen. 2008. Invited review: Crossbreeding in dairy cattle: A Danish perspective. Journal of Dairy Science 91:4116-4128.

Sorensen, L.K., M. Lund, and B. Juul. 2003. Accuracy of Fourier Transform infrared spectrometry in determination of casein in dairy cows' milk. Journal of Dairy Research 70:445-452.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D.P. Berry, M. Coffey and, P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. Journal of Dairy Science 94:1657-1667.

Soyeurt, H., P. Dardenne, F. Dehareng, C. Bastin, and N. Gengler. 2008. Genetic parameters of saturated and monounsaturated fatty acid content and the ratio of saturated to unsaturated fatty acids in bovine milk. Journal of Dairy Science 91:3611-3626.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. Journal of Dairy Science 89:3690-3695.

Stipanuk, M.H. 2004. Sulfur amino acid metabolism: pathways for production and removal of homocysteine and cysteine. Annual Revision of Nutrition 24:539-577.

Stuart, B. 2004. Infrared spectroscopy Wiley Library Online cow milk using mid-infrared spectrometry. Journal of Dairy Science 89:3690-3695.

Swinehart, D.F. 1962. The Beer-Lambert Law. Journal of Chemical Education 39:333.

Taylor, V.J., D.E. Beever, M.J. Bryant, and D.C. Wathes. 2003. Metabolic profiles and progesterone cycles in first lactation dairy cows. Theriogenology 59:1661-1677.

Teagasc. 2014. International Dairy federation. Accessed June 21st, 2015. http://www.teagasc.ie/publications/2013/1999/WorldMilkDay2013.pdf.

Teagasc. 2014. Accessed Sep. 21st, 2015. http://www.teagasc.ie/dairy/grass-basednutrition/docs/Whats_in_Grass_Todays_Farm_May2014.pdf.

Teagasc Annual Report. 2011. Accessed August 5th, 2015. http://www.teagasc.ie/publications/2012/1591/Annual_Report_2011_English.pdf.

Thaller, G., W. Krämer, A. Winter, B. Kaupe, G. Erhardt, and R. Fries. 2003. Effects of variants on milk production traits in German cattle breeds. Journal of Animal Science 81:1911-1918.

Tian, R., W.S. Pitchford, C.A. Morris, N.G. Cullen, and C.D.K. Bottema. 2010. Genetic variation in the b, b-carotene-9', 10'-dioxygenase gene and association with fat colour in bovine adipose tissue and milk. Animal Genetics 41:253-259.

Urbano-Cuadrado, M., M.D. Luque de Castro, P.M. Pérez-Juan, J. García-Olmo, and M. A. Gómez-Nieto. 2004. Near infrared reflectance spectroscopy and multivariate analysis in enology: Determination or screening of fifteen parameters in different types of wines. Analytica Chimica Acta 527:81-88.

Van Vleck, L.D., E.J. Pollak, and E.A. Oltenacu. 1987. Genetics for the animal sciences. WH Freeman and Company 207.

VanRaden, P.M., and A.H. Sanders. 2003. Economic merit of crossbred and purebred US dairy cattle. Journal of Dairy Science 86:1036-1044.

Vasbinder, A.J., and C.G. De Kruif. 2003. Casein-whey protein interactions in heated milk: the influence of pH. Internaional Dairy Journal 13:669-677.

Veerkamp RF, P. Dillon, E. Kelly, AR. Cromie, and AF. Groen. 2002. Dairy cattle breeding objectives combining yield survival and calving interval for pasture-based systems in Ireland under different milk quota scenarios. Livestock Production Science 76:137-151.

Visentin, G., A. McDermott, S. McParland, D.P. Berry, O.A. Kenny, A. Brodkorb, M.A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. Journal of Dairy Science 98:1-10.

Visser, S, C.J. Slangen, and H.S. Rollema. 1991. Phenotyping of bovine milk proteins by reversed-phase high performance liquid chromatography. Journal of Chromatography 548:361-370.

Wadhwani, R., and D.J. McMahon. 2012. Color of low-fat cheese influences flavor perception and consumer liking. Journal of Dairy Science 95:2336-2346.

Walker, G.P., C. Wijesundera, F.R. Dunshea, and P.T. Doyle. 2013. Seasonal and stage of lactation effects on milk fat composition in northern Victoria. Animal Production Science 53:560-572.

Wedholm, A., L.B. Larsen, H. Lindmark-Mansson, A.H. Karlsson, and A. Andren. 2006. Effect of Protein Composition on the Cheese-Making Properties of Milk from Individual Dairy Cows. Journal of Dairy Science 89:3296-3305.

Weikard, R., C. Kühn, T. Goldammer, G. Freyer, and M. Schwerin. 2005. The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. Physiological Genomics 21:1-13.

Welsh, F.S. 1993. The newsletter of architectural finishes investigation from Frank S. Welsh company. Fall 1:2.

Willham, R.L., and E. Pollsk. 1985. Symposium: Heterosis and Crossbreeding. Journal of Dairy Science 68:9.

Williams, P. 2007. Statistical terms for evaluation of accuracy and precision. Near infrared technology – getting the best out of light. PDK Grain, Nanaimo, BC, and Winnipeg, Man., Canada 5:1-17

Winkelman, A.M., D.L. Johnson, and A.K.H. MacGibbon. 1999. Estimation of heritabilities and correlations associated with milk colour traits. Journal of Dairy Science 82:215-224.

Wohlt, J.E., D.H. Kleyn, G.W. Vandernoot, D.J. Selfridge, and C.A. Novotney. 1981. Effect of stage of lactation, age of ewe, sibling status, and sex of lamb on gross

and minor constituents of dorset ewe milk. Journal of Dairy Science 64:2175-2184.
Wray, N., and P. Visscher. 2008. Estimating trait heritability. Nature Education 1(1):29.
Wu, M., C. Shaung, L. Gang, W. Zhifang, W. Qingjjang, H. Pingang, and F. Yuzhi. 2012. Rapid Determination of Free Amino Acids in Milk by Microchip Electrophoresis Coupled with Laser-induced Fluorescence Detection. Chinese Journal of Chemistry 30:2517-2521.
Wu, G. 2009. Amino Acids: metabolism, functions, and nutrition. Amino Acids 37:1-17
Zhang, X., B. Kong, and Y.L. Xiong. 2007. Production of cured meat color in nitrite-free Harbin red sausage by Lactobacillus fermentum fermentation. Meat Science 77:593-598.