



A landscape of the heritability of single-band Fourier-Transform Infrared spectra data in Canadian Holstein

**G. Rovere^{1,2,3,5}, G. de los Campos^{2,3,4}, R. J. Tempelman¹, A.I. Vazquez^{2,3},
F. Miglior^{6,7}, F. Schenkel⁶, A. Cecchinato⁸, G. Bittante⁸, H. Toledo-Alvarado⁸**

¹Animal Science, Michigan State University (MSU), USA

²Epidemiology and Biostatistics, Michigan State University (MSU), USA

³Institute for Quantitative Health Science and Engineering, Michigan State University (MSU), USA

⁴Statistics and Probability, Michigan State University (MSU), USA

⁵Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, Denmark

⁶Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Canada

⁷Canadian Dairy Network, Canada

⁸Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Italy

Fourier-transform mid-infrared (FTIR) milk spectra data is routinely collected within milk recording programs of different countries. This information can be used both for assessing milk composition and for genetic evaluations. Establishing an optimal strategy for the use of spectra data in genetic evaluations require knowledge of the heritability of individual bands. Therefore, in this study we used data from about 1.8 million test-day records of Canadian Holstein cows to produce a landscape of the heritability of FTIR data by band (1,060 evenly-spaced bands), parity (from first to third) and month of the lactation (from 1st to 4th). Several regions of the spectrum that have been reported to be associated to important milk components (e.g., lactose, fat and protein) showed moderate-to-high heritability estimates (0.40-0.50). We confirm many of the heritability patterns reported in previous studies and report novel findings related to differences in the heritability of FTIR spectra across parities and month of the lactation.

Keywords: milk spectra, FTIR, spectrometry, high-throughput phenotyping, Bayesian, BGLR.

Fourier-transform infrared spectroscopy (FTIR) can be used to describe the molecular structure of many different materials. In the dairy industry, FTIR data derived from electromagnetic wave absorbance within the mid infrared (MIR) spectrum, is routinely used to assess milk composition for payment, quality control, herd management purposes as well as for selective breeding. Indeed, FTIR-derived predictions of protein (PROT%) and fat percentage (FAT%) from individual samples of milking cows are routinely stored in national recording systems and used as inputs for the genetic

Summary

Introduction

Corresponding Author: roveregabriel@gmail.com

evaluations (ICAR, 2012). The clear majority of research efforts and industry applications using FTIR data has focused on developing prediction equations for economically important phenotypes such as milk composition (Rutten *et al.*, 2010; 2011) and technological properties (Cecchinato *et al.*, 2009). FTIR data has also been shown to be a valuable tool for assessing health and reproductive phenotypes (Heuer *et al.*, 2001; Belay *et al.*, 2017) as well as feed efficiency and methane emission (Shetty *et al.*, 2017a; b).

Spectra-derived predictions of phenotypes (e.g., FAT%, PROT%) are routinely used as “traits” in genetic evaluations. However, the best predictor of a phenotype is not necessarily the best selection index for the same trait. Developing an optimal selection index requires knowledge of genetic and environmental variance and co-variance components (Hazel, 1943). A few authors have estimated and reported genetic parameters of FTIR data (e.g., Soyeurt *et al.* 2010; Bittante and Cecchinato, 2013; Wang *et al.* 2016). These studies have shown that absorbance (and transmittance) at certain regions of the spectrum can be moderately heritable.

However, the studies published so far were based on limited sample size (<2,000 cows) and did not consider the possibility that the heritability of FTIR data may vary across parities and over the course of lactation. Therefore, our goal was to produce a landscape of the heritability of FTIR spectra data across parities and over the course of the lactation. To this end we analyzed a very large Canadian dataset comprising ~1.8 million records of FTIR data from Holstein milking cows linked to covariates and pedigree data.

Materials and methods

Data (n=1:781,005 test-day records, from 560,131 Holstein cows, milked in 6,368 herds distributed in seven Canadian provinces) were from the Canadian Dairy Network (Guelph, Ontario) collected between January 2013 to June 2016. Milk samples were processed using two FTIR MilkoScan FT6000 (Foss Electric A/S, Hillerød, Denmark) spectrophotometers at the Canadian DHI organizations CanWest DHI (Guelph, Ontario) and Valacta (Sainte-Anne-de-Bellevue, Québec). Milk spectra consisted of FTIR absorbance data at 1,060 bands located between wavenumbers 5,011 cm⁻¹ and 925 cm⁻¹. In addition to FTIR spectra data, milk yield (MY) and spectra-derived predictions of PROT% and FAT% were available for each test-day milk sample. Additional information included the herd, year and test-date at which the samples were collected and a pedigree that traced back three generations from each of the individuals with phenotypic records.

Stratification. We divided data into disjoint groups defined by 3 parities (1, 2 and 3) and 4 stages of lactation as defined by 30 days in milk (DIM) intervals between calving and 120 DIM. Thus, there were a total of 12 strata. The number of records per stratum ranged from 39,164 to 82,367 (*Table 1*).

Models

We estimated the heritability of single-band absorbance spectra data, single test-day MY, PROT% and FAT% using a standard mixed model of the form:

$$y_{ij} = \mu + hys_i + a_i + e_{ij},$$

where y_i ($i=1, \dots, n$) represents a phenotype (e.g., absorbance data on one of the bands), μ represent the overall mean, hys_i is the random effect of i^{th} herd-year-season level (four seasons were considered: January-March, April-June, July-September and

Table 1. Average (SD) of milk yield (kg), and [number of records] by parity and stage of lactation.

Parity	Stage of lactation			
	1-30 DIM	31-60 DIM	61-90 DIM	91-120 DIM
1	28.7 (6.3)	32.8 (6.2)	32.8 (6.1)	31.9 (6.1)
	[71,140]	[82,367]	[81,608]	[80,795]
2	38.4 (8.2)	42.1 (8.1)	40.5 (7.8)	38.4 (7.6)
	[55,954]	[63,956]	[62,902]	[62,396]
3	40.3 (8.9)	44.7 (8.8)	43.4 (8.4)	41.0 (8.2)
	[39,164]	[44,169]	[42,985]	[42,343]

October-December, 33,119 hys levels), a_j is the additive genetic effect of j^{th} cow and ε_{ij} is an error term. The vector of herd-year-season (hys) and additive effects (a) and the vector of error terms (ε) were assumed to follow multivariate normal distributions of the form: $\text{hys} \sim \text{MVN}(0, I\sigma_{\text{hys}}^2)$, $a \sim \text{MVN}(0, A\sigma_a^2)$ and $\varepsilon \sim \text{MVN}(0, I\sigma_\varepsilon^2)$, respectively. Here I is an identity matrix, A is the pedigree relationship matrix, σ_{hys}^2 is the variance of the random combined effect of the herd-year-season factor, σ_a^2 is the additive genetic variance, and σ_ε^2 is the error variance. We implemented statistical analyses within a fully Bayesian framework.

All the analyses were carried out in the R-environment (R Core Team, 2016). The pedigree was processed using the pedigreeR R-package (Vazquez and Bates, 2017). Models were fitted using the BGLR software (Pérez and de los Campos, 2014). For each analysis, we computed posterior means, posterior standard deviations and 95% credibility regions for each of the variance parameters and for the heritability (h^2) defined as:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{\text{hys}}^2 + \sigma_\varepsilon^2}$$

where σ_a^2 , σ_{hys}^2 , and σ_ε^2 are the variance terms defined before.

The estimated heritabilities of MY, PROT% and FAT% increased from the first to the second month of the lactation (Figure 2) and then remained stable (or varied with no clear pattern) towards the third and fourth month of the lactation. Among those three traits, PROT% had the highest estimated heritability (~0.4 in the 2nd, 3rd and 4th month of the lactation), followed by FAT% (~0.25-0.30 in the same months) and MY (~0.15-0.20 for months 2-4). There were no clear differences in heritability estimates across lactations.

Figure 1 shows a landscape of the estimated h^2 of the 1,060 bands of milk spectra along the four stages of lactation for each parity. The estimated heritability curves per strata, and including 95% confidence regions are given in Figure 3. Heritability estimates ranged between values close to zero to values about 0.55. Estimated heritability curves were rather smooth, in that heritability estimates of adjacent bands for the same strata

Software

Results

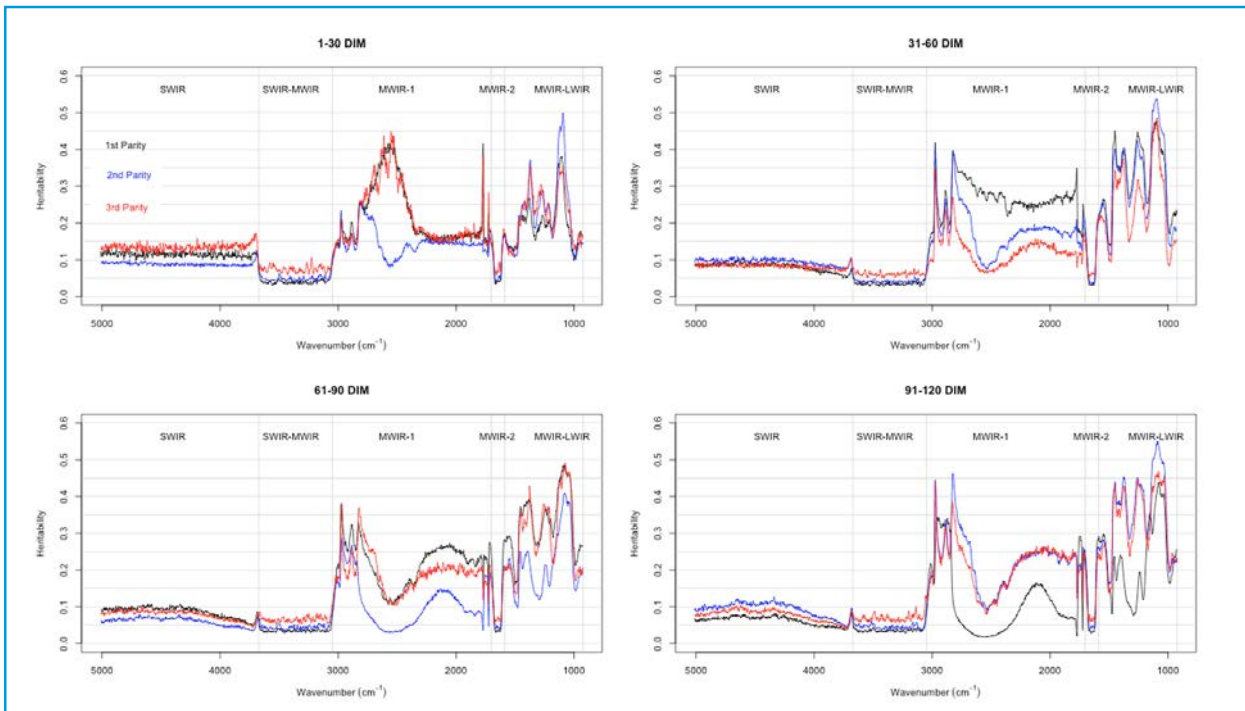


Figure 1. Heritability estimates by band (horizontal axis), parity (each color correspond to one parity) and stage of lactation (DIM).

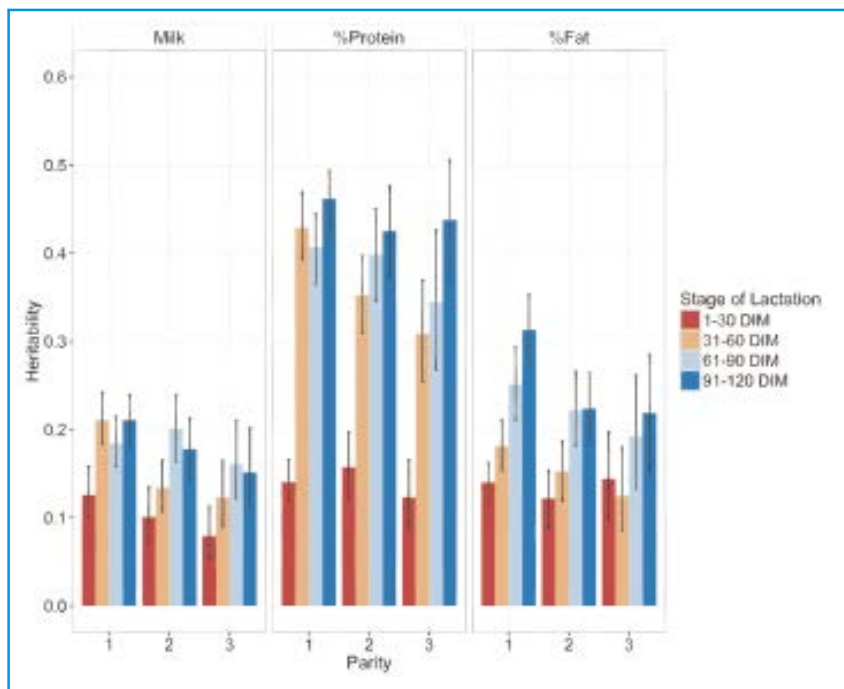


Figure 2. Heritability estimates for test-day milk yield, percentage of protein and percentage of fat by parity and stage of lactation.

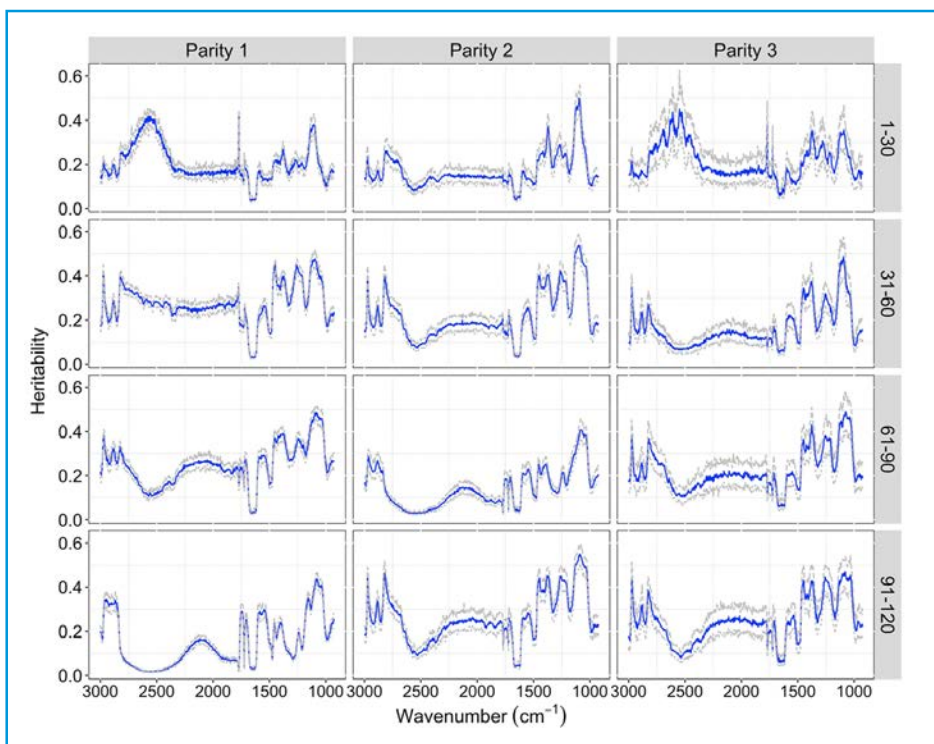


Figure 3. Estimated heritability (95% Bayesian credibility region) of individual bands in the regions between wavenumbers 3,000 cm^{-1} and wavenumbers 925 cm^{-1} , by parity and stage of the lactation.

were very similar. There was not a clear trend that indicate higher h^2 for a specific parity. However, we recognized some patterns across regions of the spectra consistent across the various strata.

Firstly, the short-wave-length infrared (**SWIR**) region of spectra (spanning between wave numbers 5,066 cm^{-1} to 3,672 cm^{-1}) and the transition region between SWIR and mid-wave-length infrared (**MWIR**, spanning from wavenumbers 3,672 cm^{-1} to 3,050 cm^{-1}) were the regions with lowest h^2 estimates, with values ranging from values near zero to 0.13. These patterns were very similar across parities and month-in-milk (Figure 1).

Secondly, the region of **MWIR** (spanning between wavenumbers 3,050 cm^{-1} to 1701 cm^{-1}) showed substantial variability in h^2 estimates across months of the lactation and parity. Here estimates ranged from near-zero to moderately high (values near 0.44). In this sector, there appeared to be two different patterns across lactation stages. For 61-90 DIM and 91-120 DIM there were two peaks of heritability: one from wavenumbers 3,050 cm^{-1} to ~ 2,800 cm^{-1} , and one with a peak of heritability estimate in the neighborhood of wavenumber 2,200 cm^{-1} . Between these two peaks is a “valley” of low-heritability bands. These patterns are clear in 61-90 DIM and 91-120 DIM for all parities (Figure 1). However, the first month of lactation (1-31 DIM) demonstrated a different pattern; here, in the first and third lactations there was a clear peak of heritability estimates around wavelength 2,700 cm^{-1} which was not present in later stages of the lactation. The results for the second month of lactation show a somehow transitional pattern between that of the first month and those observed in the third and fourth month of lactation.

Finally, the last sector of the spectra, which spans between wavenumbers 1,586 cm^{-1} and 925 cm^{-1} and comprises the **MWIR** and **long-wavelength infrared bands (LWIR)** showed heritability estimates ranging from 0.14 to 0.55. In this region, there were three consecutive peaks of high-heritability estimates with valleys in between; the pattern appeared with different intensities in all the strata analyzed.

Discussion

Our heritability estimates for test-day milk yield, FAT% and PROT% are in line with previously reported estimates for the same population (Loker *et al.*, 2009; Nixon *et al.*, 2009; Koeck *et al.*, 2013). For individual bands, we obtained heritability estimates ranging from values near zero to values of about 0.55. As noted by Bittante and Cecchinato (2013), there are multiple regions of the spectra showing higher heritability than that of yield traits commonly used in genetic evaluations. We confirmed this in the Canadian-Holstein population, using a significantly larger sample size.

The patterns of heritability estimates by regions of the spectra reported here confirm those reported by previous authors for other populations (Bittante and Cecchinato, 2013; Soyeurt *et al.*, 2010). However, we found substantially higher heritability estimates in many regions of the spectra. Several factors may explain this. Beside the differences in the breed-lines and production systems considered, it is worth noting that our analyses were stratified by parity and month of the lactation and that we consider only the first 120 DIM; therefore, we did not treat spectra from different stages of the lactation curve as the same trait.

The comparison of our results with those previously reported for different regions of the spectra leads to some important conclusions. Firstly, similar to what was reported by Bittante and Cecchinato (2013) we found that the SWIR, SWIR-MWIR and MWIR-2 are regions characterized by low-heritability estimates. Secondly, the MWIR-LWIR regions have some of the wavelengths with highest heritability. The patterns of heritability found in these regions (three waves of high heritability estimates) are very consistent across parities and lactation stage; however, these patterns are not as well defined in the first month of lactation. This region is known to be related to important milk components (e.g., lactose and some proteins). Finally, in the MWIR region the patterns of heritability estimates of the 3rd and 4th month of lactation were different than those of the 1st month of the lactation. Interestingly, in the first month of lactation the patterns of heritability estimates for 2nd parity was different to those of the first parity. We do not have a clear explanation of what may explain these differences in heritability between early and advance stages of lactations; however, considering the molecules that have been link to the bands on this region (e.g. N-H ammonium ions) it is possible that the different patterns observed may be linked to energy-balance-related aspects, this hypothesis needs to be further examined.

Acknowledgments

GR, AIV, RT acknowledge financial support from the USDA (PAPM-EAGER, 2017-67007-25947). The development and maintenance of the BGLR package was supported by NIH (R01 101219). Funding for collection of milk spectra is by Agriculture and Agri-Food Canada and by additional contributions from Dairy Farmers of Canada, the Canadian Dairy Network and the Canadian Dairy Commission under the Agri-Science Clusters Initiative.

List of References

- Belay, T.K., M. Svendsen, Z.M. Kowalski, and T. Adnoy.** 2017. Genetic parameters of blood β -hydroxybutyrate predicted from milk infrared spectra and clinical ketosis, and their associations with milk production traits in Norwegian Red cows. *J. Dairy Sci.* 100:6298–6311.
- Bittante, G., and A. Cecchinato.** 2013. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. *J. Dairy Sci.* 96:5991–6006.
- Cecchinato, A., M. De Marchi, L. Gallo, G. Bittante, and P. Carnier.** 2009. Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk. *J. Dairy Sci.* 92:5304–5313.
- Ferragina, A., G. de los Campos, A.I. Vazquez, A. Cecchinato, and G. Bittante.** 2015. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *J. Dairy Sci.* 98:8133–8151.
- Hazel, L.N.** 1943. The genetic basis for constructing selection indexes. *Genetics* 28:476–490.
- Heuer, C., H.J. Luinge, E.T. Lutz, Y.H. Schukken, J.H. van der Maas, H. Wilmink, and J.P. Noordhuizen.** 2001. Determination of acetone in cow milk by Fourier transform infrared spectroscopy for the detection of subclinical ketosis. *J. Dairy Sci.* 84:575–582.
- ICAR,** 2012. Technical Series no.16. International Strategies and new developments in milk analysis. VI ICAR Reference Laboratory Network Meeting. Cork, Ireland. Editor: O. Leray (available 27/09/17 at: www.icar.org)
- Koeck, A., F. Miglior, J. Jamrozik, D.F. Kelton, and F.S. Schenkel.** 2013. Genetic associations of ketosis and displaced abomasum with milk production traits in early first lactation of Canadian Holsteins. *J. Dairy Sci.* 96:4688–4696.
- Loker, S., F. Miglior, J. Bohmanova, L.R. Schaeffer, J. Jamrozik, and G. Kistemaker.** 2009. Short communication: Effect of preadjusting test-day yields for stage of pregnancy on variance component estimation in Canadian Ayrshires. *J. Dairy Sci.* 92:2270–2275.
- Nixon, M., J. Bohmanova, J. Jamrozik, L.R. Schaeffer, K. Hand, and F. Miglior.** 2009. Genetic parameters of milking frequency and milk production traits in Canadian Holsteins milked by an automated milking system. *J. Dairy Sci.* 92:3422–3430.
- Pérez, P., and G. De Los Campos.** 2014. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495.
- R Core Team.** 2016. R: A Language and Environment for Statistical Computing.
- Rutten, M.J.M., H. Bovenhuis, and J.A.M. van Arendonk.** 2010. The effect of the number of observations used for Fourier transform infrared model calibration for bovine milk fat composition on the estimated genetic parameters of the predicted data. *J. Dairy Sci.* 93:4872–4882.
- 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. *J. Dairy Sci.* 94:5683–5690.

Shetty, N., G. Difford, J. Lassen, P. Loevendahl, and A.J. Buitenhuis. 2017a. Predicting methane emissions of lactating Danish Holstein cows using Fourier transform mid-infrared spectroscopy of milk. *J. Dairy Sci.* doi:<https://doi.org/10.3168/jds.2017-13014>.

Shetty, N., P. Loevendahl, M.S. Lund, and A.J. Buitenhuis. 2017b. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. *J. Dairy Sci.* 100:253-264.

Soyeurt, H., I. Misztal, and N. Gengler. 2010. Genetic variability of milk components based on mid-infrared spectral data. *J. Dairy Sci.* 93:1722–1728.

Vazquez, A. I. and D. M. Bates. 2017. pedigreeR (<https://github.com/Rpedigree>).

Wang, Q., A. Hulzebosch, and H. Bovenhuis. 2016. Genetic and environmental variation in bovine milk infrared spectra. *J. Dairy Sci.* 99:6793–6803.