J. Dairy Sci. 106:3397–3410
https://doi.org/10.3168/jds.2022-22351

Genome-wide association study and functional annotation analyses for nitrogen efficiency index and its composition traits in dairy cattle

Y. Chen,[1]* H. Atashi,[1,2] C. Grelet,[3] R. R. Mota,[4] S. Vanderick,[1] H. Hu,[1] GplusE Consortium,† and N. Gengler[1]
[1]TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech (ULiège-GxABT), 5030 Gembloux, Belgium
[2]Department of Animal Science, Shiraz University, 71441-65186 Shiraz, Iran
[3]Walloon Agricultural Research Center (CRA-W), 5030 Gembloux, Belgium
[4]Council on Dairy Cattle Breeding, Bowie, MD 20716

ABSTRACT

The aims of this study were (1) to identify genomic regions associated with a N efficiency index (NEI) and its composition traits and (2) to analyze the functional annotation of identified genomic regions. The NEI included N intake (NINT1), milk true protein N (MTPN1), milk urea N yield (MUNY1) in primiparous cattle, and N intake (NINT2+), milk true protein N (MTPN2+), and milk urea N yield (MUNY2+) in multiparous cattle (2 to 5 parities). The edited data included 1,043,171 records on 342,847 cows distributed in 1,931 herds. The pedigree consisted of 505,125 animals (17,797 males). Data of 565,049 SNPs were available for 6,998 animals included in the pedigree (5,251 females and 1,747 males). The SNP effects were estimated using a single-step genomic BLUP approach. The proportion of the total additive genetic variance explained by windows of 50 consecutive SNPs (with an average size of about 240 kb) was calculated. The top 3 genomic regions explaining the largest rate of the total additive genetic variance of the NEI and its composition traits were selected for candidate gene identification and quantitative trait loci (QTL) annotation. The selected genomic regions explained from 0.17% (MTPN2+) to 0.58% (NEI) of the total additive genetic variance. The largest explanatory genomic regions of NEI, NINT1, NINT2+, MTPN1, MTPN2+, MUNY1, and MUNY2+ were *Bos taurus* autosome 14 (1.52–2.09 Mb), 26 (9.24–9.66 Mb), 16 (75.41–75.51 Mb), 6 (8.73–88.92 Mb), 6 (8.73–88.92 Mb), 11 (103.26–103.41 Mb), 11 (103.26–103.41 Mb). Based on the literature, gene ontology, Kyoto Encyclopedia of Genes and Genomes, and protein-protein interaction, 16 key candidate genes were identified for NEI and its composition traits, which are mainly expressed in the milk cell, mammary, and liver tissues. The number of enriched QTL related to NEI, NINT1, NINT2+, MTPN1, and MTPN2+ were 41, 6, 4, 11, 36, 32, and 32, respectively, and most of them were related to the milk, health, and production classes. In conclusion, this study identified genomic regions associated with NEI and its composition traits, and identified key candidate genes describing the genetic mechanisms of N use efficiency-related traits. Furthermore, the NEI reflects not only its composition traits but also the interactions among them.

Key words: gene, QTL, enrichment analysis

INTRODUCTION

High-efficiency dairy cattle are increasingly being pursued by milk producers (Brito et al., 2020). Cattle consume N mostly in the form of feed crude protein, which is then degraded into different forms of N, such as amino acids, ammonia, and urea, for metabolism in the body (Aguirre-Villegas et al., 2017). The N emissions from the livestock sector account for one-third of current human-induced N emissions (Uwizeye et al., 2020). Dairy cows with high N use efficiency (**NUE**) not only improve the profitability of dairy farms but also

Received May 30, 2022.
Accepted October 24, 2022.
*Corresponding author: yansen.chen@uliege.be
†List of authors within the GplusE consortium: Mark Crowe, Alan Fahey, Fiona Carter, Elizabeth Matthews, Andreia Santoro, Colin Byrne, Pauline Rudd, Roisin O'Flaherty, Sinead Hallinan, Claire Wathes, Mazdak Salavati, Zhangrui Cheng, Ali Fouladi, Geoff Pollott, Dirk Werling, Beatriz Sanz Bernardo, Conrad Ferris, Alistair Wylie, Matt Bell, Mieke Vaneetvelde, Kristof Hermans, Miel Hostens, Geert Opsomer, Sander Moerman, Jenne De Koster, Hannes Bogaert, Jan Vandepitte, Leila Vandevelde, Bonny Vanranst, Klaus Ingvartsen, Martin Tang Sorensen, Johanna Hoglund, Susanne Dahl, Soren Ostergaard, Janne Rothmann, Mogens Krogh, Else Meyer, Leslie Foldager, Charlotte Gaillard, Jehan Ettema, Tine Rousing, Torben Larsen, Victor H. Silva de Oliveira, Cinzia Marchitelli, Federica Signorelli, Francesco Napolitano, Bianca Moioli, Alessandra Crisà, Luca Buttazzoni, Jennifer McClure, Daragh Matthews, Francis Kearney, Andrew Cromie, Matt McClure, Shujun Zhang, Xing Chen, Huanchun Chen, Junlong Zhao, Liguo Yang, Guohua Hua, Chen Tan, Guiqiang Wang, Michel Bonneau, Marlène Sciarretta, Armin Pearn, Arnold Evertson, Linda Kosten, Anders Fogh, Thomas Andersen, Matthew Lucy, Chris Elsik, Gavin Conant, Jerry Taylor, Deborah Triant, Nicolas Gengler, Michel Georges, Frederic Colinet, Marilou Ramos Pamplona, Hedi Hammami, Catherine Bastin, Haruko Takeda, Aurelie Laine, Anne-Sophie Van Laere, Rodrigo Mota, Saeid Naderi Darbagshahi, Frederic Dehareng, Clement Grelet, Amelie Vanlierde, Eric Froidmont, Frank Becker, Martin Schulze, Sergio Palma Vera.

reduce environmental N pollution (Calsamiglia et al., 2010). The NUE is a complex trait involving multiple features, such as N intake (**NINT**), milk true protein N (**MTPN**), and MUN (Chen et al., 2022). Milk urea concentration (**MU**) and MUN are the most commonly used NUE proxies in dairy cattle management and genetic breeding programs. The reason why MU (MUN) indirectly increases NUE is its strong correlation with urinary N (Kauffman and St-Pierre, 2001).

The traditional definition of NUE in dairy cows is milk N out divided by NINT. However, several shortcomings of this definition were shown in our latest study (Chen et al., 2022). Therefore, we proposed a new N efficiency index (**NEI**) that considers both NUE and N pollution at the same time (Chen et al., 2022). The NEI is a combination of NINT, MTPN, and MUN yield (**MUNY**) being predicted by milk mid-infrared (**MIR**) spectroscopy. The genetic correlations between NEI and production yield traits were positive, but the correlations with the investigated functional traits were negative (Chen et al., 2022). However, the biological background of NEI is still missing. Although some studies have explained the biological background of MUNY or MUN (Strucken et al., 2012; Ariyarathne et al., 2021; Honerlagen et al., 2021), to our best knowledge, the biological backgrounds of NINT and MTPN have yet not been investigated. In addition, some studies performed genetic analyses on minor N compounds in milk (such as ammonia) and urinary urea (Pegolo et al., 2018; Honerlagen et al., 2021). However, these phenotypes are difficult to measure and thus far have been too challenging to be applied for dairy breeding purposes.

Multiple studies have shown that single-step genome-wide association study (**ssGWAS**) is an efficient method for studying the genomic background of complex traits (Li et al., 2019; Atashi et al., 2020; Brunes et al., 2021). Indeed, the ssGWAS algorithm can directly obtain the SNP variance through the genomic EBV, allowing estimation of the proportion of each SNP in the total additive genetic variance (Wang et al., 2012). However, the variance effect of a single SNP is often small, so it is a good way to express the proportion of genomic regions (SNP windows) of several consecutive SNPs in the total additive genetic variance (Fragomeni et al., 2014). The functional analysis of genes inside the identified genomic regions can better explain the genomic background of the research traits. For example, the gene ontology (**GO**) and Kyoto Encyclopedia of Genes and Genomes (**KEGG**) enrichment analyses of genes located in genomic regions associated with a trait can reveal the biological process and pathways involved. Based on the genetic relationships between NEI and other traits at different strength levels (Chen et al., 2022), we speculate that the genetic region identified for NUE-related features may also regulate other traits. In addition, previous studies reported that the QTL of MUNY were located on different chromosomes, showing the polygenic profile of this trait (Bouwman et al., 2010; Strucken et al., 2012). Identified genomic regions can be compared with the QTL (genomic regions) previously reported and checked for the potential effects of genetic selection of NUE on other traits at the QTL level.

The objectives of this study were to investigate the genomic background of the NEI and verify whether the NEI can reflect the combined 3 NUE-related features. In this regard, ssGWAS was used to identify genomic regions associated with NEI and its composition traits; then, functional annotation analyses were performed on the genomic regions identified for the corresponding traits.

MATERIALS AND METHODS

The study framework is shown in Figure 1. Because no human or animal subjects were used, this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

Data

Phenotypic Data. The data used in this study were the same as those used by Chen et al. (2022). Briefly, we used 1,043,171 test-day records, collected from 2001 to 2019 on 342,847 cows distributed in 1,931 herds. The range of DIM for all records used was restricted to between 5 and 50 because of the predicament of the NINT model. The NINT of each cow was predicted by the equations based on the models developed by Grelet et al. (2020). The following formulas were used: MTPN = [(milk yield × protein percentage/6.38) − MUNY], and MUNY = [(milk urea concentration/2.14) × milk yield] (WHO and FAO, 2011), with protein percentage and milk urea concentration predicted by milk MIR. In addition, we divided the 3 NUE-related features (NINT, MTPN, MUNY) into **NINT1**, **MTPN1**, **MUNY1**, **NINT2+**, **MTPN2+**, and **MUNY2+** traits according to primiparous or multiparous classes (2 to 5 parities). The used pedigree consisted of 505,125 animals (17,797 males). Grelet et al. (2020) used support vector machine regression to build a NINT prediction model based on the milk MIR spectra, milk yield, and parity. A total of 143,595 NINT records were predicted with data from 53,660 cows in 776 herds.

Genotypic Data. Genotypic data were available for 6,998 animals (1,747 males and 5,251 females). Individ-

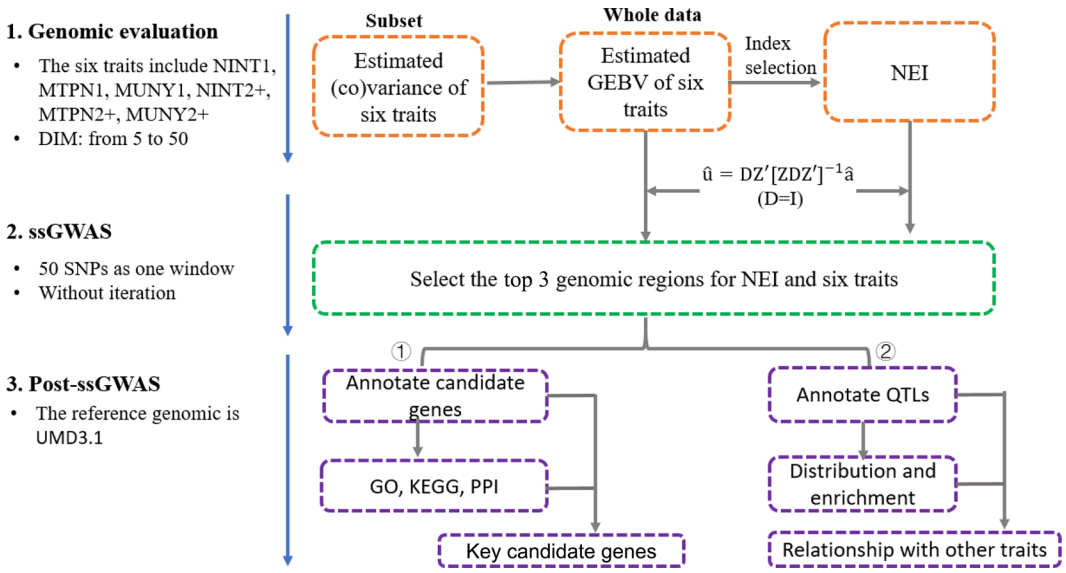

Figure 1. Workflow for the N efficiency index (NEI), N intake in primiparous cows (NINT1), N intake in multiparous cows (NINT2+), milk true protein N in primiparous cows (MTPN1), milk true protein N in multiparous cows (MTPN2+), MUN yield in primiparous cows (MUNY1), and MUN yield in multiparous cows (MUNY2+). GO = gene ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes; PPI = protein-protein interaction; ① = first showing these results in this manuscript; ② = showing these results after ①; GEBV = genomic EBV; ssGWAS = single-step GWAS.

uals were genotyped using the BovineSNP50 Beadchip v1 to v3 (Illumina). Single nucleotide polymorphisms common between all 3 chips were kept. Non-mapped SNPs, SNPs located on sex chromosomes, and triallelic SNPs were excluded. A minimum GenCall Score of 0.15 and a minimum GenTrain Score of 0.55 were used to keep SNPs (Wilmot et al., 2022). The genotypes were imputed to HD by using FImpute V2.2 software (Sargolzaei et al., 2014). One of the common editing steps for marker data (e.g., SNP) is to check for Mendelian conflicts (Wiggans et al., 2009). A Mendelian conflict occurs when the genotype and pedigree data of 2 related animals are in disagreement. This may result from an error in the recorded pedigree, from genotyping errors, from mixing up DNA samples, and, in very rare cases, from mutations (Calus et al., 2011). In this study, SNPs with Mendelian conflicts and those with minor allele frequency less than 5% were excluded. The difference between observed heterozygosity and that expected under Hardy-Weinberg equilibrium was estimated, and SNPs difference greater than 0.15 were excluded (Wiggans et al., 2009). In total, 565,049 SNPs located on 29 BTA were used in the genomic analyses.

(Co)Variance Components Estimation. In total, 143,595 test-day records on 53,660 cows for 6 traits extracted from the whole data set were used to estimate variance and covariance components (Chen et al., 2022). The pedigree used for estimating (co)variance components consisted of 133,943 animals (7,879 males). Genotypic data were available for 4,563 animals (1,292 males) included in the pedigree. A 6-trait repeatability model (6 traits: 3 traits and 2 parity classes) was used to estimate the (co)variance components. The information of the model can be found in Chen et al. (2022). In brief, herd-year-season of calving, DIM, and calving age (nested within parities) were used as fixed effects in this model, whereas nongenetic cow, nongenetic cow × parity (only for multiparous traits), additive animal genetic, and residual were used as random effects. However, when calculating the relationship between animals, we used the **H** matrix, which combined pedigree (**A**) and genomic (**G**) based relationships into one matrix. The inverse of **H** as defined by Aguilar et al. (2010) is as follows:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

where **A** is the numerator relationship matrix based on the pedigree; $\mathbf{A}_{22}$ is the numerator relationship matrix based on the pedigree for genotyped animals; **G** is the genomic relationship matrix obtained using the function described by VanRaden (2008). In addition, the inverse of all matrices considers the coefficient of inbreeding between individuals (Lourenco et al., 2020).

Computations were performed using the BLUPF90 family of programs (Misztal et al., 2018). The (co)vari-

ance components and parameters for NINT, MTPN, and MUNY were estimated by Gibbs sampling, which was described by Chen et al. (2022). The procedures used to calculate the heritability (h^2) and repeatability were the same as those reported by Chen et al. (2021).

Estimated Breeding Values and Nitrogen Efficiency Index

Using the estimated (co)variance components and the same model, genomic EBV were estimated for 6 traits of each animal, according to the precondition conjugate algorithm implemented in the BLUPF90 (version 1.71) program. The whole data set was used for this purpose. The calculation method of NEI was the same as Chen et al. (2022). In short, The NEI was obtained by combining the GEBV of the 6 traits using the selection index theory. The relative weights of the 6 considered traits were calculated by selection responses, which assumed that the selection responses for NINT, MTPN, and MUNY were 0, 1, and −1, respectively.

Genome-Wide Association Analyses

The SNP effects for the NEI and 6 traits were estimated using the POSTGSF90 software (version 1.73; Aguilar et al., 2014). The formula used for estimating SNP effects was as follows Wang et al. (2012):

$$\hat{\mathbf{a}} = \mathbf{D}\mathbf{Z}_{\boldsymbol{g}}'\left[\mathbf{Z}_{\mathbf{g}}\mathbf{D}\mathbf{Z}_{\boldsymbol{g}}'\right]^{-1}\hat{\mathbf{u}},$$

where $\hat{\mathbf{a}}$ is the SNP effect, $\mathbf{D}$ is the weight matrix of SNPs, which is identical to the identity matrix ($\mathbf{I}$), which means the weight for all SNPs is 1; $\mathbf{Z}_{\mathbf{g}}$ is an incidence matrix of genotyped for each SNP; and $\hat{\mathbf{u}}$ is a vector of genomic EBV for each trait of genotyped animals. The variance of i SNP is $\hat{d}_i = a_i^2 2p_i\left(1-p_i\right)$, where a_i^2 is the square of ith SNP effect, and p_i is the frequency of allele B at SNP i. The results were presented by the proportion of variance explained by each window of 50 adjacent SNPs with an average size of about 240 kb. We used 1 SNP as the moving step of the window, which ensured that we would not miss genomic regions potentially associated with the trait due to the combination of SNPs. The formula for the total additive genetic variance of each window was as follows:

$$\frac{var\left(a_i\right)}{\sigma_a^2}\times 100(\%) = \frac{var\left(\sum_{j=1}^{50}\mathbf{Z}_{\mathbf{j}}\hat{\mathbf{d}}_{\mathbf{j}}\right)}{\sigma_a^2}\times 100(\%),$$

where a_i is the genetic variance of the ith genomic region (each window combines 50 consecutive SNPs), σ_a^2 is the total genetic variance, $\mathbf{Z}_{\mathbf{j}}$ is the vector of the SNP content of the jth SNP for all individuals, and $\hat{\mathbf{d}}_{\mathbf{j}}$ is the variance of the jth SNP.

Linkage disequilibrium (squared correlation coefficient, r^2) was calculated for SNPs within a window that explained more than 0.5% of total additive genetic variance.

Functional Annotation Analyses

Following Soares et al. (2021), the top genomic regions were selected to investigate candidate genes and their annotation. However, due to the large number of traits considered and the small proportion of variance explained by genomic regions in the current study, only the top 3 genomic regions were selected. Then, candidate genes and QTL annotations were performed through the GALLO R package (Fonseca et al., 2020).

Protein-encoding genes located in these selected genomic regions were identified using the *Bos taurus* UMD3.1.94 assembly as the reference map (http://ftp.ensembl.org/pub/release-94/gtf/bos_taurus/; accessed Oct. 19, 2021). The GO and KEGG analyses were carried out on the identified candidate gene sets obtained for NEI and included 6 traits through the clusterProfiler R package (Wu et al., 2021). Furthermore, protein-protein interaction (**PPI**) analysis was performed on the candidate genes obtained from the analyzed traits through STRING (Szklarczyk et al., 2021) to reveal the relationships between the identified candidate genes. The PPI relationship was based on text mining, experiments, database, co-expression, neighborhood, gene fusion, and co-occurrence, and the minimum required interaction score was set to 0.40 (Zhou et al., 2019). The Cytohubba in Cytoscape (version 3.8.2) was used to find the hub genes. Based on the literature, GO, KEGG, and PPI, potential candidate genes were selected, hereafter referred to as key candidate genes. Moreover, we checked the expression levels of the identified candidate genes (or key genes) over 100 tissues or cell types in cattle through the cGTEx database (https://cgtex.roslin.ed.ac.uk/; Liu et al., 2022).

The top 3 genomic regions identified for the studied traits were annotated with Cattle QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/BT/index; accessed Oct. 19, 2021; Hu et al., 2019). At present, Cattle QTLdb has 158,041 QTL, which were divided into 6 classes including exterior, production, health, reproduction, milk, meat, and carcass (https://www.animalgenome.org/cgi-bin/QTLdb/BT/ontrait?class_ID = 1). To avoid deviation caused by the annota-

Table 1. Heritability, repeatability, σ_a^2, σ_c^2, σ_p^2 (only for second and later lactations), and σ_e^2 of the proxies for predicted NINT, MTPN, and MUNY in primiparous and multiparous Holstein cows (mean ± SE)[1]

Trait[2]	h^2	Repeatability	σ_a^2	σ_c^2	σ_p^2	σ_e^2
NINT1	0.13 ± 0.01	0.38 ± 0.01	0.03 ± 0.00[3]	0.06 ± 0.00	NA[4]	0.14 ± 0.00
MTPN1	0.12 ± 0.00	0.59 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	NA	0.02 ± 0.00
MUNY1	0.14 ± 0.01	0.40 ± 0.01	0.10 ± 0.01	0.20 ± 0.01	NA	0.45 ± 0.01
NINT2+	0.12 ± 0.01	0.45 ± 0.00	0.04 ± 0.00	0.03 ± 0.00	0.08 ± 0.00	0.17 ± 0.00
MTPN2+	0.10 ± 0.01	0.64 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.03 ± 0.00	0.03 ± 0.00
MUNY2+	0.10 ± 0.01	0.43 ± 0.00	0.16 ± 0.01	0.13 ± 0.01	0.40 ± 0.01	0.93 ± 0.01

[1]σ_a^2 = additive genetic variance; σ_c^2 = across-parity permanent environment (nongenetic cow) variance; σ_p^2 = within-parity permanent environment (nongenetic cow × parity) variance; σ_e^2 = residual variance; NINT = predicted N intake, expressed as 100 g/d; MTPN = milk true protein N, expressed as 100 g/d; MUNY = MUN yield, expressed as g/d; primiparous cows, n = 44,321; multiparous cows, n = 99,374.

[2]Traits: NINT1 = N intake in primiparous cows; MTPN1 = milk true protein N in primiparous cows; MUNY1 = MUN yield in primiparous cows; NINT2 = N intake in multiparous cows; MTPN2 = milk true protein N in multiparous cows; MUNY2 = MUN yield in multiparous cows.

[3]SE < 0.005.

[4]NA = not applicable.

tion richness of the different traits, the hypergeometric test approach was adopted for the enrichment analyses (Fonseca et al., 2020). In all enrichment analyses (GO, KEGG, QTL), the Benjamini-Hochberg method was used for multiple testing corrections.

RESULTS AND DISCUSSION

Genetic Parameters

The genetic parameters estimated for the 6 considered traits are described in Table 1, and are similar to our previous results obtained without using genotype data (Chen et al., 2022). The h^2 and repeatability for the 6 considered traits ranged from 0.10 to 0.14, and 0.38 to 0.64, respectively. Compared with our previous study (Chen et al., 2022), the changes in variances ranged from −3.00% to 0.88%; the absolute value of h^2 changed from 4.64% (MUNY1) to 8.81% (MUNY2+), and that of repeatability changed from 0.36% (NINT2+) to 3.32% (NINT1). Therefore, it can be concluded that including genotypic data in the variance components analysis caused a minor effect on the results, and the 6 considered traits have low h^2 and medium repeatability, although we expected that using genotype data would capture more variance and increase the h^2, especially for complex traits. One possible reason is that the number of genotyped animals in this work was small (n < 5,000; de los Campos et al., 2018); hence, substantial changes were not observed.

Genome-Wide Association Analyses

Manhattan plots for the NEI and its composition traits are shown in Figure 2. The top 3 genomic regions selected for the NEI and its composition traits are shown in Table 2. The genomic regions identified for NEI, NINT, MTPN, and MUNY were in BTA 11, 14; 8, 16, 22, 25, and 26; 6, 14, 13, 18, and 19; and 6, 8, and 11, respectively (Table 2). The genomic regions identified for NEI were also associated with MTPN and MUNY (except the third region), which confirmed that NEI is associated with MTPN and MUNY. The identified genomic regions explained from 0.17% (MTPN2+) to 0.58% (NEI) of the total additive genetic variance. Only NEI has genomic regions that explain more than 0.50% of the total additive genetic variance, probably because NEI responds to its component trait interactions.

The top 3 genomic regions combined explained 1.18, 0.75, 0.73, 0.58, 0.59, 0.96, and 1.03% of the total additive genetic variance for NEI, NINT1, NINT2+, MTPN1, MTPN2+, MUNY1, and MUNY2+, respectively. Results showed that, in general, windows explained less than 0.50% of the total additive genetic variance of the traits, and these low-contributing regions were spread across the genome for all traits analyzed. This indicates that NEI and its composition traits are moderate to highly polygenic, in which many regions across the genome contribute to the genetic variation of the traits. Similar results were reported for MUNY by Strucken et al. (2012). It should be noted that one window explaining more than 0.50% of the total additive genetic variance was identified only for NEI, which means that NEI may reflect interactive effects between MTPN and MUNY. The linkage disequilibrium estimated for the genomic region with more than 0.50% of the total additive genetic variance of NEI is shown in Figure 3. We found that 38.89% (7/18) of the genes involved in the nitrogen metabolism pathway (KEGG: 00910) were located in BTA14 (27.63 to 79.73 Mb) and included *CA1*, *CA2*, *CA3*, *CA8*, *CA13*, *LOC784254*, and *LOC10084787*4. The BTA14 position between 27.63 and 79.73 Mb is near the top 1 genomic

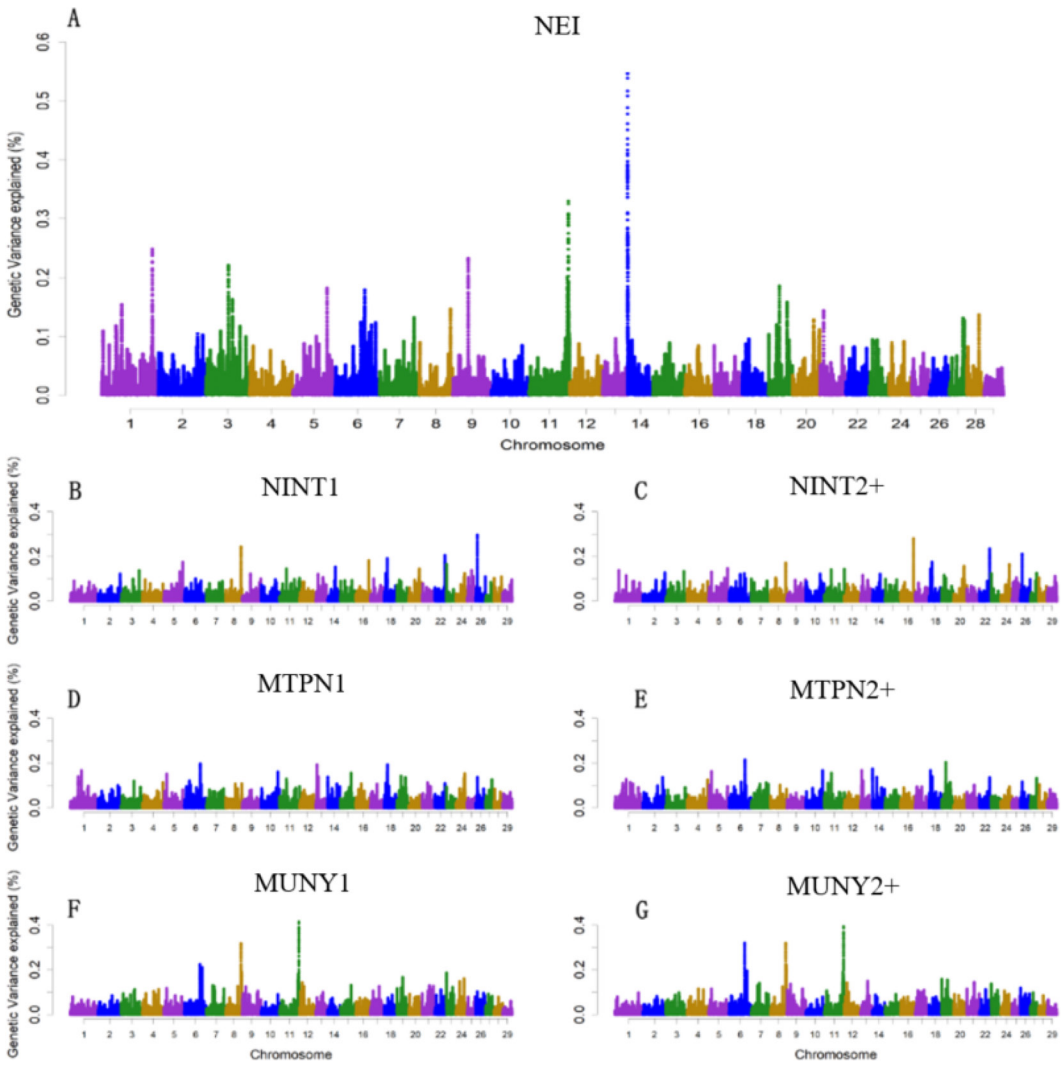

Figure 2. Total additive genetic variance is explained by windows of 50 adjacent SNP across chromosomes for the N efficiency index (NEI, A), the N intake in primiparous cows (NINT1, B), N intake in multiparous cows (NINT2+, C), milk true protein N in primiparous cows (MTPN1, D), milk true protein N in multiparous cows (MTPN2+, E), MUN yield in primiparous cows (MUNY1, F), and MUN yield in multiparous cows (MUNY2+, G).

region (BTA14, 1.54 to 2.09 Mb) related to NEI. From this point, only the genomic regions that explained the largest additive genetic variance of the studied traits were discussed.

The genomic region located between 1.52 and 2.09 Mb on BTA14 explained 0.58% of the total additive genetic variance of NEI. Multiple studies have reported that this region is associated with milk yield (Nayeri et al., 2016; Atashi et al., 2020; Bakhshalizadeh et al., 2021). This region (BTA14: 1.52 to 2.09 Mb) was also associated with protein and fat yields in dairy cows (Veerkamp et al., 2016; Cai et al., 2019). This is in line with our prediction of improving NEI through improving MTPN. The genomic regions between 9.24 to 9.66 Mb on BTA26 and between 75.41 and 75.51 Mb on BTA16 explained the largest part of the total additive genetic variance of NINT1 and NINT2+, respectively. Previous studies have shown that SNPs inside the region found on BTA26 position 9.24 to 9.66 Mb were associated with milk yield, milk C14 index, and milk myristoleic acid content (Minozzi et al., 2013; Gebreyesus et al., 2019). The genomic region identified at BTA16 position 75.41 to 75.51 Mb was associated with residual feed intake and feed efficiency in cattle (Brunes et al., 2021). The genomic region between 88.73 and 88.92 Mb on BTA6 accounted for the largest ratio of total additive genetic variance for MTPN1 and MTPN2+. This region has been reported to be associated with protein yield and composition (Olsen et al., 2016; Zhou et al., 2019). The genomic region located between 103.26 and 103.41 Mb

Table 2. Annotated genes within the top 3 genomic regions explain the largest proportion of total genetic variance for nitrogen efficiency index (NEI) and its composition traits

Trait	BTA	Position[2] (bp)	Var[3] (%)	Gene[4]
NEI	14	1517553:2089613	0.58	*ZNF7*, *COMMD5*, *ARHGAP39*, *C14H8orf82*, *LRRC24*, *LRRC14*, *RECQL4*, *MFSD3*, *GPT*, *PPP1R16A*, *FOXH1*, *KIFC2*, *CYHR1*, *TONSL*, *VPS28*, *SLC39A4*, *CPSF1*, *ADCK5*, *SLC52A2*, *FBXL6*, *TMEM249*, *SCRT1*, *DGAT1*, *HSF1*, *BOP1*, *SCX*, *MROH1*, *HGH1*, *WDR97*, *MAF1*, *SHARPIN*, *CYC1*, *GPAA1*, *EXOSC4*, *OPLAH*, *SPATC1*, *GRINA*, *PARP10*, *PLEC*
	11	103264921:103409247	0.33	*PAEP*, *GLT6D1*, *LCN9*, *KCNT1*
	14	2673388:2978629	0.27	*GML*, *LY6K*, *LY6D*, *LYNX1*, *LYPD2*, *SLURP1*, *THEM6*, *PSCA*, *ARC*, *ADGRB1*
NINT1	26	9242669:9655433	0.30	*PAPSS2*, *ATAD1*, *PTEN*
	8	103696313:103829659	0.24	*SNX30*, *SLC46A2*
	22	55490915:55638639	0.20	*SLC6A11*
NINT2+	16	75405390:75509546	0.28	*IRF6*, *C16H1orf74*, *TRAF3IP3*, *HSD11B1*
	22	55490915:55638639	0.23	*SLC6A11*
	26	9242669:9655433	0.21	*PAPSS2*, *ATAD1*, *PTEN*
MTPN1	6	88732184:88919352	0.20	*GC*
	13	10175391:10315354	0.19	*KIF16B*
	18	15797080:15884324	0.19	*ITFG1*
MTPN2+	6	88732184:88919352	0.22	*GC*
	19	22594096:22657020	0.20	*NXN*, *MRM3*, *GLOD4*, *DBIL5*
	14	1517553:2089613	0.17	*ZNF7*, *COMMD5*, *ARHGAP39*, *C14H8orf82*, *LRRC24*, *LRRC14*, *RECQL4*, *MFSD3*, *GPT*, *PPP1R16A*, *FOXH1*, *KIFC2*, *CYHR1*, *TONSL*, *VPS28*, *SLC39A4*, *CPSF1*, *ADCK5*, *SLC52A2*, *FBXL6*, *TMEM249*, *SCRT1*, *DGAT1*, *HSF1*, *BOP1*, *SCX*, *MROH1*, *HGH1*, *WDR97*, *MAF1*, *SHARPIN*, *CYC1*, *GPAA1*, *EXOSC4*, *OPLAH*, *SPATC1*, *GRINA*, *PARP10*, *PLEC*
MUNY1	11	103264921:103409247	0.41	*PAEP*, *GLT6D1*, *LCN9*, *KCNT1*
	8	103694244:103828116	0.32	*INIP*, *SNX30*, *SLC46A2*
	6	87136725:87296185	0.22	*CSN1S1*, *CSN2*, *HSTN*, *STATH*, *CSN1S2*
MUNY2+	11	103264921:103409247	0.39	*PAEP*, *GLT6D1*, *LCN9*, *KCNT1*
	6	87145250:87311202	0.32	*CSN1S1*, *CSN2*, *HSTN*, *STATH*, *CSN1S2*
	8	103696313:103829659	0.32	*SNX30*, *SLC46A2*

[1]Traits: NINT1 = N intake in primiparous cows; MTPN1 = milk true protein N in primiparous cows; MUNY1 = MUN yield in primiparous cows; NINT2 = N intake in multiparous cows; MTPN2 = milk true protein N in multiparous cows; MUNY2 = MUN yield in multiparous cows.

[2]Starting and ending coordinates of the genomic region.

[3]Var = percentage of genetic variance explained by the SNPs within the genomic region.

[4]Genes: EBSEMBL symbol of annotated genes using the *Bos Taurus* UMD 3.1.94 assembly (http://ftp.ensembl.org/pub/release-94/gtf/bos_taurus/).

on BTA11 was associated with MUNY1 and MUNY2+. This region has been reported to be associated with MU in Brown Swiss cattle (Pegolo et al., 2018) and MUN in Australian and New Zealand dairy cattle (van den Berg et al., 2022a). Ariyarathne et al. (2021), using the 50K SNP chip, found that the region located between 100 and 101 Mb on BTA11 was associated with MU in mixed-breed cattle (Holstein Friesian, Jersey, and Holstein Friesian × Jersey crossbred). Previous studies have shown that the genomic region located between 87 and 88 Mb on BTA6 was associated with MU (Pegolo et al., 2018; Ariyarathne et al., 2021). This region was among the identified top 3 genomic regions associated with MUNY. Previous studies have reported that the region between 6.12 and 7.15 Mb on BTA17 was associated with MU yield (Honerlagen et al., 2021). The variation observed in different studies can be explained by the number of animals used in their studies.

Briefly, the genomic regions identified for the studied traits were located on multiple BTA and explained a small fraction of the total additive genetic variance, suggesting that these are complex quantitative traits controlled by multiple genes. Increasing the SNP density in these genomic regions of NEI and 6 considered traits (especially NINT and MUNY) when making the SNP chip may improve the reliability of genetic selection for NUE. The NEI may better reflect NUE because it has a prominent peak at BTA14, which is closer to genes related to nitrogen metabolism pathways.

Gene Annotation Analyses

The results of gene annotation analyses are described in Table 2 and Figure 4. We detected no common candidate genes between NEI and NINT (Figure 4), which is consistent with our hypothesis (keeping NINT unchanged). Surprisingly, we found no common candidate gene between NEI and MTPN1, which could be because NEI only increases the NUE of dairy cows through MTPN2+ in the first 5 parities. The percent-

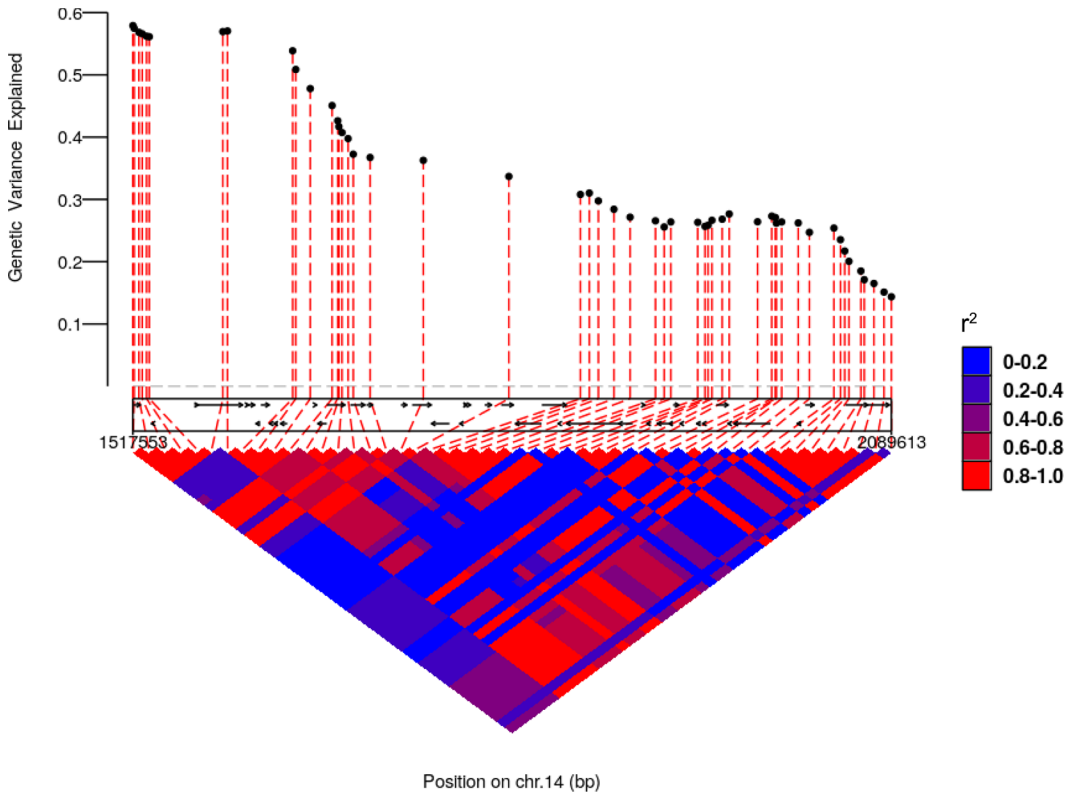

Figure 3. Linkage disequilibrium between 50 SNPs inside the genomic region on BTA14, position 1.52 to 2.09 Mb, associated with nitrogen efficiency index. Chr. = chromosome.

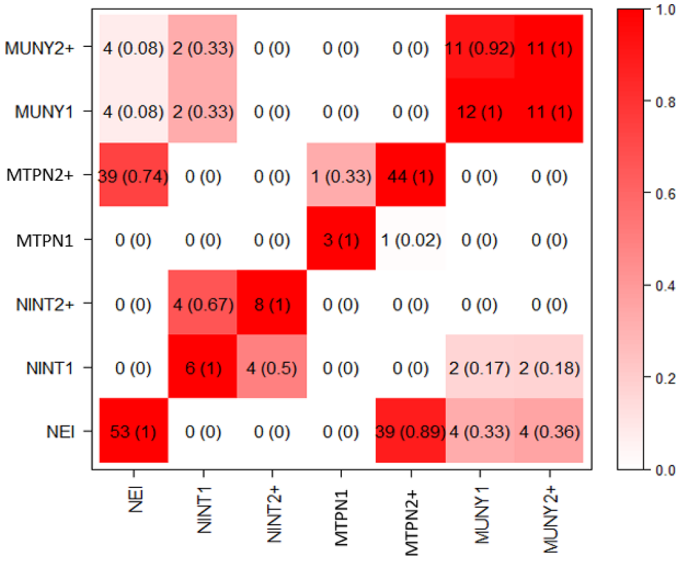

Figure 4. Numbers (percentage) of shared genes among N efficiency index (NEI), N intake in primiparous cows (NINT1), N intake in multiparous cows (NINT2+), milk true protein N in primiparous cows (MTPN1), milk true protein N in multiparous cows (MTPN2+), MUN yield in primiparous cows (MUNY1), and MUN yield in multiparous cows (MUNY2+). The upper left triangle is the number of candidate genes for NEI as the denominator, and the lower right triangle is the number of candidate genes for traits as the denominator.

age of common candidate genes for NEI and MTPN2+, MUNY1, MUNY2+ ranged from 33 to 89%.

Among the annotated candidate genes, *DGAT1*, *GRINA*, *CYHR1*, *FOXH1*, *TONSL*, *PPP1R16A*, *ARHGAP39*, *MAF1*, *OPLAH*, *MROH1*, *ZNF7*, *SLURP1*, *MAFA*, *KIFC2*, *GML*, *PSCA*, *THEM6*, *LYNX1*, and *ARC* have been reported to be associated with 305-d milk yield (Nayeri et al., 2016; Atashi, et al., 2020). The *DGAT1* gene has also been reported to be importantly associated with milk yield, fat, and protein percentages (Bakhshalizadeh et al., 2021). The *CSN1S1*, *CSN1S2*, *CSN2*, and *PAEP* genes have been reported to be associated with milk protein composition (Sanchez et al., 2017; Pegolo et al., 2018; Zhou et al., 2019). The *BOP1* gene has been associated with protein yield (Cai et al., 2019). Brunes et al. (2021) reported that *MAF1* was associated with low animal feed intake.

The results of the GO enrichment analysis are presented in Supplemental File S1 (https://github.com/Yansen0515/NEI_GWAS_POST_GWAS; Chen, 2022). The candidate genes identified for NINT1, NINT2+, MTPN1, MUNY1, and MUNY2+ enriched 33, 5, 3, 23, and 28 GO terms, respectively; however, the candidate genes identified for NEI and MTPN2+ enriched no GO

terms. The 33 GO terms enriched for NINT involve only *ATAD1*, which was also identified as a candidate gene for DMI (Serão et al., 2013). The MTPN1 trait obtained the 3 GO terms by *GC*, which encode the vitamin D binding protein. The *GC* gene has been identified as being associated with milk production, mastitis, and postpartum blood calcium concentration (Olsen et al., 2016; Cavani et al., 2022). Of the 28 GO terms enriched for MUNY through *CSN1S1*, *CSN1S2*, *CSN2*, *HSTN*, and *STATH*, the first 3 genes (*CSN1S1*, *CSN1S2*, *CSN2*) belong to CSN@ (casein cluster) family genes, which have been identified as being related to α_{S1}-CN, β-CN, and κ-CN (Zhou et al., 2019). *HSTN* has been identified as being related to κ-CN (Zhou et al., 2019), as well as affecting β-CN and α_{S2}-CN (Bovine Genome Sequencing and Analysis Consortium, 2009). *STATH* affects β-CN (Rijnkels et al., 2003).

Only MUNY1 and MUNY2 enriched 2 pathways: the salivary secretion (bta04970) and prolactin signaling pathways (bta04917; Supplemental File S2; https://github.com/Yansen0515/NEI_GWAS_POST_GWAS; Chen, 2022). These pathways were enriched by *CSN2*, *HSTN*, and *STATH*, and the 3 genes were explained the same as in the GO analysis part. Salivary secretion affects feed intake in cattle (Taussat et al., 2020). Salivary secretion also is associated with methane emissions in nonlactating dairy cows (Pinares-Patiño et al., 2007). The prolactin signal regulates the milk production and composition of dairy cows (Raven et al., 2014).

The internal gene network of the candidate gene sets for NEI and its composition traits are shown separately in Supplemental File S3 (from Supplemental Figures S1 to S5; https://github.com/Yansen0515/NEI_GWAS_POST_GWAS; Chen, 2022). The PPI for NEI, NINT1, NINT2+, MTPN1, MTPN2+, MUNY1, and MUNY 2+ were composed of 58 nodes and 188 edges, 11 nodes and 8 edges, 13 nodes and 9 edges, 8 nodes and 15 edges, 49 nodes and 156 edges, 17 nodes and 32 edges, and 16 nodes and 41 edges, respectively. The PPI enrichment *P*-values of NEI, MTPN, and MUNY were less than $4 \times 1.0E^{-4}$. The NEI had a similar PPI to MTPN2+, which was also similar to the PPI of dairy components (Bakhshalizadeh et al., 2021). This is possibly caused by the candidate genes annotated in BTA14 positions 1.52 to 2.09 Mb. Both the PPI of MUNY1 and MUNY2+ showed the protein network in the STRING database (CL:24892), which was composed of *CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, and *PAEP*. *PAEP* encodes β-LG, and the first 4 genes encode casein. These 5 genes have been subjected to long-term selection and their SNP frequencies in cattle altered (Kolenda and Sitkowska, 2021). The top 1 hub genes for NEI, NINT2+, MTPN1, MTPN2+, MUNY1, and MUNY2+ were respectively *MROH1*, *PTEN*, *AHSG*, *MROH1*, *CSN1S1*, and *CSN2*. However, the PPI of NINT1 did not have hub genes. *MROH1*, *CSN1S1*, and *CSN2* affect milk protein composition (Sanchez et al., 2017). The activation of *PTEN* was not conducive to lactation of dairy cows and reduced the production of β-CN (Wang et al., 2014).

In short, NEI and MTPN were affected by genes associated with milk yield and components, such as *DGAT1*, *PPP1R16A*, *CYHR1*, *CPSF1*, *MROH1*, *GC*, and *AHSG*. The NINT trait was affected by the *ATAD1* and *PTEN* genes. The MUNY trait was affected by genes that control the production of casein (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*, *HSTN*, *STATH*, and *PAEP*) and was related to salivary secretion (bta04970). The expressions of the 16 key candidate genes involved in the cattle tissues are given in Supplemental File S4 (https://github.com/Yansen0515/NEI_GWAS_POST_GWAS; Chen, 2022), and these genes are mainly expressed in the milk cell, mammary, and liver tissues. The results of these genes and their high-expression tissue locations can be used for future studies on the genetic mechanisms of NUE.

QTL Annotation for Select Genomic Regions

The numbers of previously reported QTL located in the identified genomic regions for the studied traits are described in Supplemental File S5 (Supplemental Figure S6; https://github.com/Yansen0515/NEI_GWAS_POST_GWAS; Chen, 2022). The proportions of selected QTL in the QTL classification and the top 10 QTL in the milk class are shown in Supplemental File S5 (Figures S7 to S13; https://github.com/Yansen0515/NEI_GWAS_POST_GWAS; Chen, 2022). However, different traits have had different numbers of annotation studies, which creates a bias for annotation results (Fonseca et al., 2020). We performed enrichment analyses separately on NEI and the composition traits based on annotated QTL. The numbers and classes of significantly related QTL are shown in Figures 5 and 6, respectively. The percentages of common significant QTL between NEI and MTPN1, MTPN2+, MUNY1, and MUNY2+ were 3, 33, 11, and 12%, respectively, which is also similar to our results from constructing NEI and annotating genes. The significant QTL of NEI and MTPN were distributed only in the milk, health, and production classes, except that MTPN1 had one QTL in the exterior class. If the NEI was used in genetic selection, it means that NEI will affect these 3 trait classes, which is consistent with the genetic correlations found between NEI and the traits (Chen et al., 2022). The significant QTL annotated by NINT

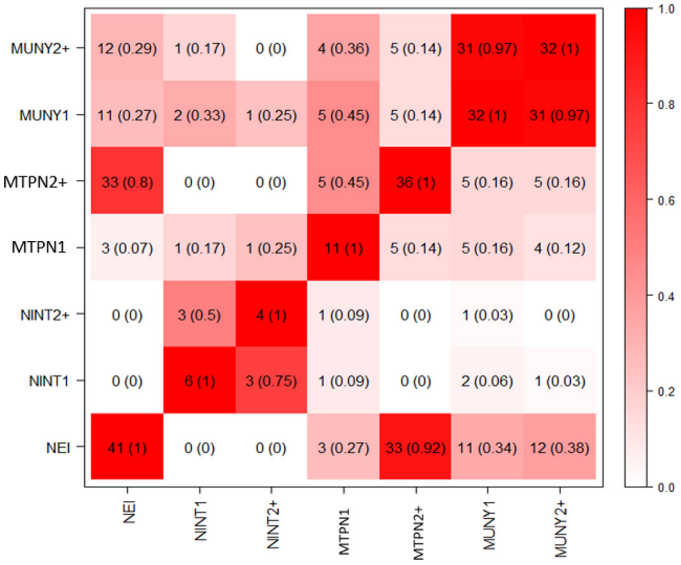

Figure 5. Numbers (percent) of enrichment QTL among N efficiency index (NEI), N intake in primiparous cows (NINT1), N intake in multiparous cows (NINT2+), milk true protein N in primiparous cows (MTPN1), milk true protein N in multiparous cows (MTPN2+), MUN yield in primiparous cows (MUNY1), and MUN yield in multiparous cows (MUNY2+). The upper left triangle is the number of candidate genes for NEI as the denominator, and the lower right triangle is the number of candidate genes for traits as the denominator.

were distributed in the milk and production classes. The significant QTL annotated by MUNY were distributed into 4 classes of QTL, which may indicate the complexity of MUNY.

The top 10 QTL after QTL enrichment analysis of the study's traits are shown in Supplemental File S6 (from Supplemental Figures S14 to S20; https://github.com/Yansen0515/NEI_GWAS_POST_GWAS; Chen, 2022). The most significant related QTL for NEI and MTPN2+ were those reported for milk yield, which is located on BTA14. This genomic region, located at BTA14 position 1.52 to 2.09 Mb, has been reported to be associated with milk production (Nayeri et al., 2016; Atashi et al., 2020; Bakhshalizadeh et al., 2021). We also found that the lifetime profit index was related to the NEI, which can be explained by the fact that the lifetime profit index is related to the BTA14 1.6 to 1.8 Mb genomic region (Nayeri et al., 2017). The most significant QTL for MTPN1 was also associated with clinical mastitis and is located at BTA6 position 88.73 to 88.92 Mb (Olsen et al., 2016; Freebern et al., 2020). The most significant QTL for NINT1 and NINT2+ has also been associated with the milk C14 index. The C14 index has been found to be associated with the genomic region located between 14.9 and 24.9 Mb on BTA26 (Li et al., 2015), similar to the NINT trait (BTA26 position 9.24 to 9.66 Mb). The most significant correlation QTL for MUNY1 and MUNY2+ was milk β-LG percentage, which may be because BTA11 position 103.26 to 103.41 Mb has been reported to be related to milk β-LG percentage (Sanchez et al., 2017). The region located at BTA11 position 103.26 to 103.41 Mb has also been reported to be associated with MU and milk protein components (Pegolo et al., 2018; Ariyarathne et al., 2021). Other studies have reported that the QTL of MUNY are located on BTA3, BTA6, and BTA21 (Bouwman et al., 2010; Strucken et al., 2012).

In summary, the related QTL of the NEI were mainly reported for milk yield, fat, and protein composition. The related QTL of NINT was reported for the milk C14 index. The related QTL of MTPN1 and MTPN2+ were related to clinical mastitis and milk yield, respectively. The related QTL of MUNY were reported for protein composition.

The traits in this study were all related to traits predicted by milk MIR. The effects of using traits predicted from milk MIR have disadvantages and advantages for their influence on GWAS results. The disadvantage is that, as with other predictive traits, the accuracy of the prediction equation can have a large effect. If the accuracy of the prediction equation is very low, GWAS may not work on the traits we are interested in. The advantage of milk MIR prediction traits is that they can be predicted at low costs and on a large scale. Large-scale data are useful to overcome the problem of inaccurate GWAS results from small samples. Recently, van den Berg et al. (2022b) showed that using the blood urea nitrogen predicted by milk MIR increased the power of GWAS results. In total, we believe that the use of predictive traits with high accuracy facilitates the results of GWAS. In addition, there was no effect of weighted single-step genomic BLUP on the ssGWAS results of this study under the condition of 50K chip data (detailed results not shown), which is consistent with other recent studies (Aguilar et al., 2019; Cesarani et al., 2021). One of the reasons for this result is that no SNP had a large effect on the traits studied.

CONCLUSIONS

This study explained the genomic background of NEI and its composition traits that can be used in dairy cattle breeding. The 16 key candidate genes influencing the genetic mechanism of NUE-related traits were identified; these are mainly expressed in the milk cell, mammary, and liver tissues. The NEI reflects not only the 6 studied NUE-related traits but also the interactions between them, because only the NEI has a prominent peak at BTA14, explaining more than 0.50% of the total additive genetic variance. Furthermore, the

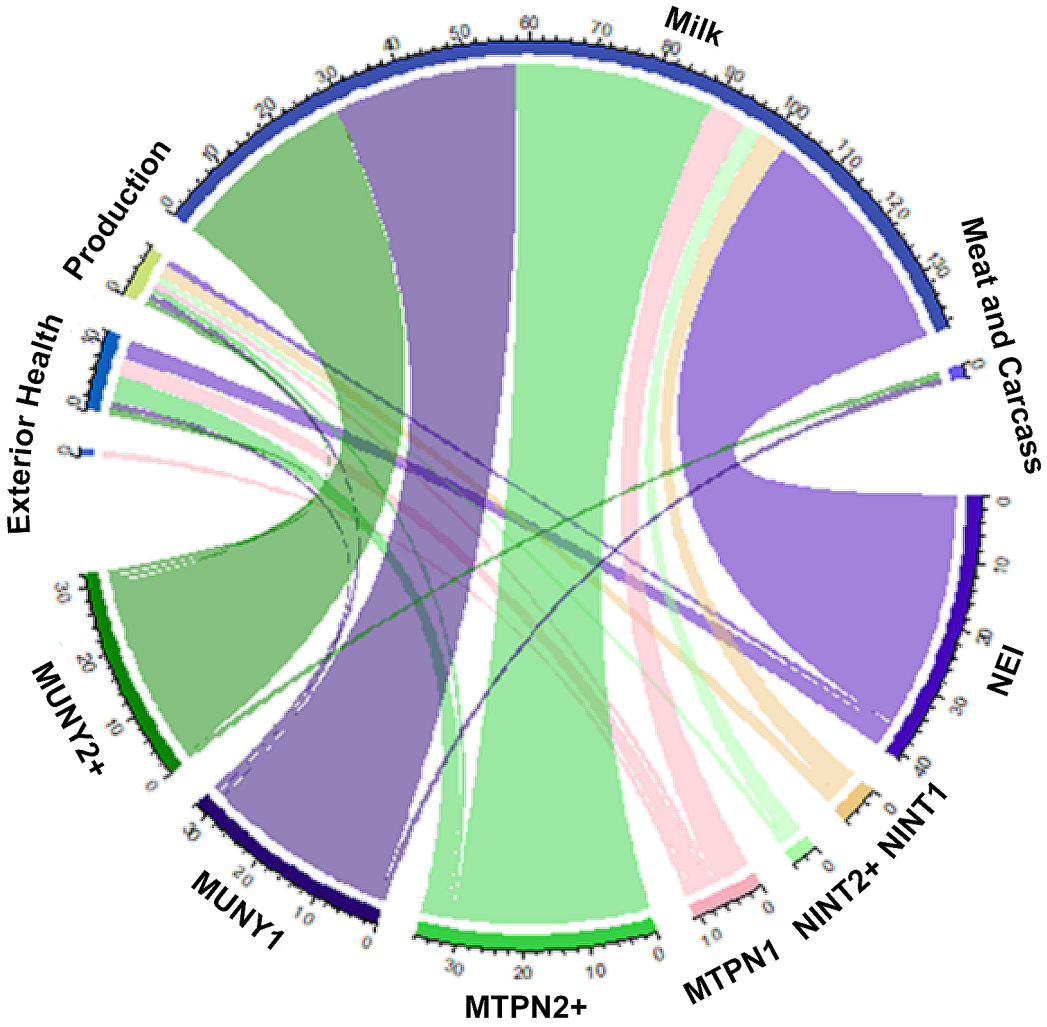

Figure 6. Numbers of enrichment QTL classes among N efficiency index (NEI), N intake in primiparous cows (NINT1), N intake in multiparous cows (NINT2+), milk true protein N in primiparous cows (MTPN1), milk true protein N in multiparous cows (MTPN2+), MUN yield in primiparous cows (MUNY1), and MUN yield in multiparous cows (MUNY2+). Quantitative trait loci classes were defined by the cattle QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/BT/ontrait?class_ID = 1).

NEI may be more representative of NUE, because the genomic regions most associated with it are closer to genes in the nitrogen metabolism pathway.

ACKNOWLEDGMENTS

The China Scholarship Council (Beijing) is acknowledged for funding the PhD project of Yansen Chen. We thank the European Union's Seventh Framework Program (Brussels, Belgium) for funding the GplusE project under grant agreement no. 613689. The views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. We thank Stefano Biffani for giving good suggestions for this work (IBBA-CNR National Research Council, Milano, Italia). We acknowledge the support of the Walloon Government (Service Public de Wallonie—Direction Générale Opérationnelle Agriculture, Ressources Naturelles et Environnement, SPW-DGARNE; Namur, Belgium), and the use of the computation resources of the University of Liège–Gembloux Agro-Bio Tech (Gembloux, Belgium) provided by the technical platform Calcul et Modélisation Informatique (CAMI) of the TERRA Teaching and Research Centre, partly supported by the National Fund for Scientific Research (FRS-FNRS, Brussels, Belgium) under grants no. T.0095.19 (PDR "DEEPSELECT") and J.0174.18 (CDR "PREDICT-2"). The authors have not stated any conflicts of interest.

REFERENCES

Aguilar, I., A. Legarra, F. Cardoso, Y. Masuda, D. Lourenco, and I. Misztal. 2019. Frequentist *P*-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. Genet. Sel. Evol. 51:28. https://doi.org/10.1186/s12711-019-0469-3.

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. https://doi.org/10.3168/jds.2009-2730.

Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. Proc. 10th World Congress of Genetics Applied to Livestock Production. Accessed Aug. 21, 2022. http://www.ainfo.inia.uy/digital/bitstream/item/15445/1/Aguilar-et-al.-2014.-WCGALP.pdf.

Aguirre-Villegas, H. A., M. A. Wattiauxand, R. A. Larson, L. Chase, S. D. Ranathunga, and M. D. Ruark. 2017. Dairy cow nitrogen efficiency. University of Wisconsin–Extension. Accessed May 28, 2022. https://www.virtualfarm.psu.edu/assets/uploads/content/DairyCap_DairyCowNitrogenEfficiency-Final-Final.pdf.

Ariyarathne, H. B. P. C., M. Correa-Luna, H. T. Blair, D. J. Garrick, and N. Lopez-Villalobos. 2021. Identification of genomic regions associated with concentrations of milk fat, protein, urea and efficiency of crude protein utilization in grazing dairy cows. Genes (Basel) 12:456. https://doi.org/10.3390/genes12030456.

Atashi, H., M. Salavati, J. De Koster, J. Ehrlich, M. Crowe, G. Opsomer, the GplusE Consortium, and M. Hostens. 2020. Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. J. Anim. Breed. Genet. 137:292–304. https://doi.org/10.1111/jbg.12442.

Bakhshalizadeh, S., S. Zerehdaran, and A. Javadmanesh. 2021. Meta-analysis of genome-wide association studies and gene networks analysis for milk production traits in Holstein cows. Livest. Sci. 250:104605–14. https://doi.org/10.1016/j.livsci.2021.104605.

Bouwman, A. C., G. C. B. Schopen, H. Bovenhuis, M. H. P. W. Visker, and J. A. M. van Arendonk. 2010. Genome-wide scan to detect quantitative trait loci for milk urea nitrogen in Dutch Holstein-Friesian cows. J. Dairy Sci. 93:3310–3319. https://doi.org/10.3168/jds.2009-2829.

Bovine Genome Sequencing and Analysis Consortium. 2009. The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science 324:522–528. https://doi.org/10.1126/science.1169588.

Brito, L. F., H. R. Oliveira, K. Houlahan, P. A. S. Fonseca, S. Lam, A. M. Butty, D. J. Seymour, G. Vargas, T. C. S. Chud, F. F. Silva, C. F. Baes, A. Cánovas, F. Miglior, and F. S. Schenkel. 2020. Genetic mechanisms underlying feed utilization and implementation of genomic selection for improved feed efficiency in dairy cattle. Can. J. Anim. Sci. 100:587–604. https://doi.org/10.1139/cjas-2019-0193.

Brunes, L. C., F. Baldi, F. B. Lopes, R. B. Lôbo, R. Espigolan, M. F. O. Costa, N. B. Stafuzza, and C. U. Magnabosco. 2021. Weighted single-step genome-wide association study and pathway analyses for feed efficiency traits in Nellore cattle. J. Anim. Breed. Genet. 138:23–44. https://doi.org/10.1111/jbg.12496.

Cai, Z., B. Guldbrandtsen, M. S. Lund, and G. Sahana. 2019. Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle. BMC Genet. 20:15. https://doi.org/10.1186/s12863-019-0717-0.

Calsamiglia, S., A. Ferret, C. K. Reynolds, N. B. Kristensen, and A. M. Van Vuuren. 2010. Strategies for optimizing nitrogen use by ruminants. Animal 4:1184–1196. https://doi.org/10.1017/S1751731110000911.

Calus, M. P., H. A. Mulder, and J. W. Bastiaansen. 2011. Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genet. Sel. Evol. 43:34. https://doi.org/10.1186/1297-9686-43-34.

Cavani, L., M. B. Poindexter, C. D. Nelson, J. E. P. Santos, and F. Peñagaricano. 2022. Gene mapping, gene-set analysis, and genomic prediction of postpartum blood calcium in Holstein cows. J. Dairy Sci. 105:525–534. https://doi.org/10.3168/jds.2021-20872.

Cesarani, A., A. Garcia, J. Hidalgo, L. Degano, D. Vicario, N. P. P. Macciotta, and D. Lourenco. 2021. Genomic information allows for more accurate breeding values for milkability in dual-purpose Italian Simmental cattle. J. Dairy Sci. 104:5719–5727. https://doi.org/10.3168/jds.2020-19838.

Chen, Y. 2022. Supplemental Files 1–6. Git Hub. https://github.com/Yansen0515/NEI_GWAS_POST_GWAS.

Chen, Y., H. Atashi, C. Grelet, S. Vanderick, H. Hu, and N. Gengler. 2022. Defining a nitrogen efficiency index in Holstein cows and assessing its potential impact on the breeding program of bulls. J. Dairy Sci. 105:7575–7587. https://doi.org/10.3168/jds.2021-21681.

Chen, Y., S. Vanderick, R. R. Mota, C. Grelet, GplusE Consortium, and N. Gengler. 2021. Estimation of genetic parameters for predicted nitrogen use efficiency and losses in early lactation of Holstein cows. J. Dairy Sci. 104:4413–4423. https://doi.org/10.3168/jds.2020-18849.

de los Campos, G., A. I. Vazquez, S. Hsu, and L. Lello. 2018. Complex-trait prediction in the era of big data. Trends Genet. 34:746–754. https://doi.org/10.1016/j.tig.2018.07.004.

Fonseca, P. A. S., A. Suárez-Vega, G. Marras, and Á. Cánovas. 2020. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. Gigascience 9:giaa149. https://doi.org/10.1093/gigascience/giaa149.

Fragomeni, B. de O., I. Misztal, D. L. Lourenco, I. Aguilar, R. Okimoto, and W. M. Muir. 2014. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. Front. Genet. 5:332. https://doi.org/10.3389/fgene.2014.00332.

Freebern, E., D. J. A. Santos, L. Fang, J. Jiang, K. L. Parker Gaddis, G. E. Liu, P. M. Vanraden, C. Maltecca, J. B. Cole, and L. Ma. 2020. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC Genomics 21:41. https://doi.org/10.1186/s12864-020-6461-z.

Gebreyesus, G., A. J. Buitenhuis, N. A. Poulsen, M. H. P. W. Visker, Q. Zhang, H. J. F. van Valenberg, D. Sun, and H. Bovenhuis. 2019. Combining multi-population datasets for joint genome-wide association and meta-analyses: The case of bovine milk fat composition traits. J. Dairy Sci. 102:11124–11141. https://doi.org/10.3168/jds.2019-16676.

Grelet, C., E. Froidmont, L. Foldager, M. Salavati, M. Hostens, C. P. Ferris, K. L. Ingvartsen, M. A. Crowe, M. T. Sorensen, J. A. Fernandez Pierna, A. Vanlierde, N. Gengler, GplusE Consortium, and F. Dehareng. 2020. Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. J. Dairy Sci. 103:4435–4445. https://doi.org/10.3168/jds.2019-17910.

Honerlagen, H., H. Reyer, M. Oster, S. Ponsuksili, N. Trakooljul, B. Kuhla, N. Reinsch, and K. Wimmers. 2021. Identification of genomic regions influencing N-metabolism and N-excretion in lactating Holstein-Friesians. Front. Genet. 12:699550. https://doi.org/10.3389/fgene.2021.699550.

Hu, Z. L., C. A. Park, and J. M. Reecy. 2019. Building a livestock genetic and genomic information knowledge base through integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Res. 47(D1):D701–D710. https://doi.org/10.1093/nar/gky1084.

Kauffman, A. J., and N. R. St-Pierre. 2001. The relationship of milk urea nitrogen to urine nitrogen excretion in Holstein and Jersey cows. J. Dairy Sci. 84:2284–2294. https://doi.org/10.3168/jds.S0022-0302(01)74675-9.

Kolenda, M., and B. Sitkowska. 2021. The polymorphism in various milk protein genes in Polish Holstein-Friesian dairy cattle. Animals (Basel) 11:389. https://doi.org/10.3390/ani11020389.

Li, B., L. Fang, D. J. Null, J. L. Hutchison, E. E. Connor, P. M. VanRaden, M. J. VandeHaar, R. J. Tempelman, K. A. Weigel, and J. B. Cole. 2019. High-density genome-wide association study for residual feed intake in Holstein dairy cattle. J. Dairy Sci. 102:11067–11080. https://doi.org/10.3168/jds.2019-16645.

Li, X., A. J. Buitenhuis, M. S. Lund, C. Li, D. Sun, Q. Zhang, N. A. Poulsen, and G. Su. 2015. Joint genome-wide association study for milk fatty acid traits in Chinese and Danish Holstein populations. J. Dairy Sci. 98:8152–8163. https://doi.org/10.3168/jds.2015-9383.

Liu, S., Y. Gao, O. Canela-Xandri, S. Wang, Y. Yu, W. Cai, B. Li, R. Xiang, A. J. Chamberlain, E. Pairo-Castineira, K. D'Mellow, K. Rawlik, C. Xia, Y. Yao, P. Navarro, D. Rocha, X. Li, Z. Yan, C. Li, B. D. Rosen, C. P. Van Tassell, P. M. Vanraden, S. Zhang, L. Ma, J. B. Cole, G. E. Liu, A. Tenesa, and L. Fang. 2022. A multi-tissue atlas of regulatory variants in cattle. Nat. Genet. 54:1438–1447. https://doi.org/10.1038/s41588-022-01153-5.

Lourenco, D., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguilar, and I. Misztal. 2020. Single-step genomic evaluations from theory to practice: Using SNP chips and sequence data in blupf90. Genes (Basel) 11:790. https://doi.org/10.3390/genes11070790.

Minozzi, G., E. L. Nicolazzi, A. Stella, S. Biffani, R. Negrini, B. Lazzari, P. Ajmone-Marsan, and J. L. Williams. 2013. Genome wide analysis of fertility and production traits in Italian Holstein cattle. PLoS One 8:e80219. https://doi.org/10.1371/journal.pone.0080219.

Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica. 2018. Manual for BLUPF90 Family Programs. University of Georgia. Accessed May 28, 2022. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75. https://doi.org/10.1186/s12863-016-0386-1.

Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, S. Miller, F. Schenkel, S. S. Moore, and P. Stothard. 2017. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. J. Dairy Sci. 100:1246–1258. https://doi.org/10.3168/jds.2016-11770.

Olsen, H. G., T. M. Knutsen, A. M. Lewandowska-Sabat, H. Grove, T. Nome, M. Svendsen, M. Arnyasi, M. Sodeland, K. K. Sundsaasen, S. R. Dahl, B. Heringstad, H. H. Hansen, I. Olsaker, M. P. Kent, and S. Lien. 2016. Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. Genet. Sel. Evol. 48:79. https://doi.org/10.1186/s12711-016-0257-2.

Pegolo, S., N. Mach, Y. Ramayo-Caldas, S. Schiavon, G. Bittante, and A. Cecchinato. 2018. Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. Sci. Rep. 8:566. https://doi.org/10.1038/s41598-017-18916-4.

Pinares-Patiño, C. S., G. C. Waghorn, A. Machmüller, B. Vlaming, G. Molano, A. Cavanagh, and H. Clark. 2007. Methane emissions and digestive physiology of non-lactating dairy cows fed pasture forage. Can. J. Anim. Sci. 87:601–613. https://doi.org/10.4141/CJAS06023.

Raven, L. A., B. G. Cocks, M. E. Goddard, J. E. Pryce, and B. J. Hayes. 2014. Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genet. Sel. Evol. 46:29. https://doi.org/10.1186/1297-9686-46-29.

Rijnkels, M., L. Elnitski, W. Miller, and J. M. Rosen. 2003. Multispecies comparative analysis of a mammalian-specific genomic domain encoding secretory proteins. Genomics 82:417–432. https://doi.org/10.1016/S0888-7543(03)00114-9.

Sanchez, M. P., A. Govignon-Gion, P. Croiseau, S. Fritz, C. Hozé, G. Miranda, P. Martin, A. Barbat-Leterrier, R. Letaïef, D. Rocha, M. Brochard, M. Boussaha, and D. Boichard. 2017. Within-breed and multi-breed GWAS on imputed whole-genome sequence variants reveal candidate mutations affecting milk protein composition in dairy cattle. Genet. Sel. Evol. 49:68. https://doi.org/10.1186/s12711-017-0344-z.

Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. https://doi.org/10.1186/1471-2164-15-478.

Serão, N. V. L., D. González-Peña, J. E. Beever, G. A. Bollero, B. R. Southey, D. B. Faulkner, and S. L. Rodriguez-Zas. 2013. Bivariate genome-wide association analysis of the growth and intake components of feed efficiency. PLoS One 8:e78530. https://doi.org/10.1371/journal.pone.0078530.

Soares, R. A. N., G. Vargas, T. Duffield, F. Schenkel, and E. J. Squires. 2021. Genome-wide association study and functional analyses for clinical and subclinical ketosis in Holstein cattle. J. Dairy Sci. 104:10076–10089. https://doi.org/10.3168/jds.2020-20101.

Strucken, E. M., R. H. Bortfeldt, D. J. De Koning, and G. A. Brockmann. 2012. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43:375–382. https://doi.org/10.1111/j.1365-2052.2011.02278.x.

Szklarczyk, D., A. L. Gable, K. C. Nastou, D. Lyon, R. Kirsch, S. Pyysalo, N. T. Doncheva, M. Legeay, T. Fang, P. Bork, L. J. Jensen, and C. von Mering. 2021. The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49(D1):D605–D612. https://doi.org/10.1093/nar/gkaa1074.

Taussat, S., M. Boussaha, Y. Ramayo-Caldas, P. Martin, E. Venot, G. Cantalapiedra-Hijar, C. Hozé, S. Fritz, and G. Renand. 2020. Gene networks for three feed efficiency criteria reveal shared and specific biological processes. Genet. Sel. Evol. 52:67. https://doi.org/10.1186/s12711-020-00585-z.

Uwizeye, A., I. J. M. de Boer, C. I. Opio, R. P. O. Schulte, A. Falcucci, G. Tempio, F. Teillard, F. Casu, M. Rulli, J. N. Galloway, A. Leip, J. W. Erisman, T. P. Robinson, H. Steinfeld, and P. J. Gerber. 2020. Nitrogen emissions along global livestock supply chains. Nat. Food 1:437–446. https://doi.org/10.1038/s43016-020-0113-y.

van den Berg, I., P. N. Ho, T. V. Nguyen, M. Haile-Mariam, I. M. MacLeod, P. R. Beatson, E. O'Connor, and J. E. Pryce. 2022a. GWAS and genomic prediction of milk urea nitrogen in Australian and New Zealand dairy cattle. Genet. Sel. Evol. 54:15. https://doi.org/10.1186/s12711-022-00707-9.

van den Berg, I., P. N. Ho, T. V. Nguyen, M. Haile-Mariam, T. D. W. Luke, and J. E. Pryce. 2022b. Using mid-infrared spectroscopy to increase GWAS power to detect QTL associated with blood urea nitrogen. Genet. Sel. Evol. 54:27. https://doi.org/10.1186/s12711-022-00719-5.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

Veerkamp, R. F., A. C. Bouwman, C. Schrooten, and M. P. L. Calus. 2016. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. Genet. Sel. Evol. 48:95. https://doi.org/10.1186/s12711-016-0274-1.

Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genet. Res. (Camb.) 94:73–83. https://doi.org/10.1017/S0016672312000274.

Wang, Z., X. Hou, B. Qu, J. Wang, X. Gao, and Q. Li. 2014. *Pten* regulates development and lactation in the mammary glands of dairy cows. PLoS One 9:e102118. https://doi.org/10.1371/journal.pone.0102118.

WHO (World Health Organization) and FAO (Food and Agriculture Organization of the United Nations). 2011. Codex Alimentarius: Milk and Milk Products. 2nd ed. WHO FAO.

Wiggans, G. R., T. S. Sonstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 92:3431–3436. https://doi.org/10.3168/jds.2008-1758.

Wilmot, H., J. Bormann, H. Soyeurt, X. Hubin, G. Glorieux, P. Mayeres, C. Bertozzi, and N. Gengler. 2022. Development of a genomic tool for breed assignment by comparison of different clas-

sification models: Application to three local cattle breeds. J. Anim. Breed. Genet. 139:40–61. https://doi.org/10.1111/jbg.12643.

Wu, T., E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, and G. Yu. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb.) 2:100141. https://doi.org/10.1016/j.xinn.2021.100141.

Zhou, C., C. Li, W. Cai, S. Liu, H. Yin, S. Shi, Q. Zhang, and S. Zhang. 2019. Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. Front. Genet. 10:72. https://doi.org/10.3389/fgene.2019.00072.