Estimation of genetic parameters and single-step genome-wide association studies for milk urea nitrogen in Holstein cattle

Longgang Ma,1,2 Hanpeng Luo,1 Luiz F. Brito,3 Yao Chang,1 Ziwei Chen,1 Wenqi Lou,1 Fan Zhang,1,2 Lei Wang,1,2 Gang Guo,4 and Yachun Wang1*

1National Engineering Laboratory of Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction, MARA, College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China
2College of Animal Science, Xinjiang Agricultural University, Urumqi, 830052, China
3Department of Animal Sciences, Purdue University, West Lafayette, IN 47907
4Beijing Sunlon Livestock Development Company Limited, 10029, Beijing, China

ABSTRACT

The main objectives of this study were to estimate genetic parameters for milk urea nitrogen (MUN) in Holstein cattle and to conduct a single-step (ss)GWAS to identify candidate genes associated with MUN. Phenotypic measurements from 24,435 Holstein cows were collected from March 2013 to July 2019 in 9 dairy farms located in the Beijing area, China. A total of 2,029 cows were genotyped using the Illumina 150K Bovine Bead Chip, containing 121,188 SNP. A single-trait repeatability model was used to evaluate the genetic background of MUN. We found that MUN is a trait with low heritability (0.06 ± 0.004) and repeatability (0.12). Considering similar milk production levels, a lower MUN concentration indicates higher nitrogen digestibility. The genetic correlations between MUN and milk yield, net energy concentration, fat percentage, protein percentage, and lactose percentage were positive and ranged from 0.02 to 0.26. The genetic correlation between MUN and somatic cell score (SCS) was negative (−0.18), indicating that animals with higher MUN levels tend to have lower SCS. Both ssGWAS and pathway enrichment analyses were used to explore the genetic mechanisms underlying MUN. A total of 18 SNP (located on BTA11, BTA12, BTA14, BTA17, and BTA18) were found to be significantly associated with MUN. The genes CFAP77, CAMSAP1, CACNA1B, ADGRB1, FARPI, and INTU are considered to be candidate genes for MUN. These candidate genes are associated with important biological processes such as protein and lipid metabolism and binding to specific proteins. This set of candidate genes, metabolic pathways, and their functions provide a better understanding of the genomic architecture and physiological mechanisms underlying MUN in Holstein cattle.

Key words: milk urea nitrogen, Chinese Holstein, environmental efficiency, repeatability model, single-step genome-wide association study

INTRODUCTION

Anthropogenic nitrogen emissions are a growing worldwide problem contributing to particulate matter in the atmosphere, greenhouse gas emissions, and eutrophication of soil and water, all of which are associated with negative consequences for human, animal, and environmental health (Leip et al., 2015). Through arable land activities and grazing, emissions of N and greenhouse gases caused a loss of 34% mean species abundance. Out of the agriculture-related losses, 76% were estimated to be caused by livestock, especially through feed production (Leip et al., 2015). Cattle husbandry has also been reported to account for 36% of ammonia emissions in Europe (Müller et al., 2020). Therefore, improving the N efficiency of milk production is important to reduce environmental emissions of nitrous oxide and ammonia at the farm level (Castillo et al., 2000).

Urea, a small organic molecule synthesized in the liver from ammonia resulting from the degradation of proteins and other N compounds (Parker et al., 1995), is considered as a useful indicator of dietary N utilization and of urinary N excretion (Jonker et al., 1998; Nousiainen et al., 2004). The amounts of urea in the blood, plasma, urine, and milk are related to casein percentages and energy percentage from the feed (Roseler et al., 1993). There is a relationship between urea and energy balance (DePeters and Cant, 1992) and also possibly with milk energy concentration. Milk and blood plasma urea concentrations are strongly correlated (0.88 to 0.98; Butler et al., 1996;
Broderick and Clayton, 1997). In general, milk urea is synthesized as a consequence of the imbalance between dietary N and energy in the rumen, and protein synthesis inefficiency. Milk urea N, an indicator of urinary N (Ciszuk and Gebregziabher, 1994; Jonker et al., 1998) and urinary urea N excretions (Burgos et al., 2007), is directly related to ammonia emissions (Burgos et al., 2007) and, consequently, increased environmental burden.

Various factors are known to contribute to excess urea N production, including excessive protein or N consumption. Apart from dietary factors, specific traits, such as BW, have been described to be positively associated with urinary N excretion (Kauffman and St-Pierre, 2001) or negatively correlated with MUN concentrations (Hojman et al., 2004). Under the premise that changes in MUN do not affect milk yield, selection for decreased MUN or other indicator traits of N emissions (e.g., urine urea nitrogen, BUN) could be a feasible strategy for breeding more N-efficient animals.

Cows in early lactation have a ruminal flora that is not adapted to the shift to high-protein diets after parturition. The consequential mismatch in energy and protein may lead to increased MUN in the first month of lactation (Jorritsma et al., 2003). Intensive selection for milk yield (MY) has also resulted in unfavorable genetic responses in traits related to fertility, health, longevity, and environmental sensitivity (Brito et al., 2021). The apparent relationships of MUN with N excretion in milk and urine suggest that decreased MUN could improve the health of dairy cows and decrease environmental pollution by excessive N. Using genetic and genomic selection to simultaneously improve MY and milk quality as well as other relevant traits (e.g., health, fertility, welfare) is paramount in dairy cattle breeding programs (Miglior et al., 2017; Brito et al., 2021). Previous heritability estimates of MUN have been reported to range from 0.06 to 0.44 (Wenninger and Distl, 1993; Wood et al., 2003; Mitchell et al., 2005). No clear pattern of the genetic relationship of MUN with other economically important traits in Holstein cattle has yet been determined. Presently, MUN is a trait routinely measured in many national milk recording systems around the world, and it could be an effective and noninvasive trait for improving dairy cattle productive efficiency. Therefore, MUN and related traits (e.g., blood urea N, urine urea N, milk-related proteins, N use efficiency), under the condition that those or corresponding proxies can be easily predicted (e.g., using infrared spectroscopy), may be useful indicator traits for genetically selecting individuals with reduced N emissions.

Single-step (ss)GWAS (Wang et al., 2012), which is based on the single-step genomic BLUP approach (ssGBLUP; Aguilar et al., 2010), simultaneously integrates all the recorded phenotype, pedigree, and genomic information, leading to more accurate results. In this context, ssGWAS could be useful for understanding the underlying biological mechanisms of MUN. Therefore, the present study aimed to explore the genetic basis of MUN in Chinese Holstein cattle. We estimated genomic-based genetic parameters for MUN and other milk-related traits and mapped genomic regions associated with MUN.

**MATERIALS AND METHODS**

Ethical review and approval were not required for the animal study because all phenotypic data were recorded as part of routine dairy cattle management and genetic evaluations. The DNA samples were obtained for the purpose of routine genomic evaluations in previous projects. Thus, no additional animal handling or experiment was performed specifically for this study.

**Phenotype, Pedigree, and Genotypic Data**

The MUN records from 24,435 Holstein cows were obtained from March 2013 to July 2019 from 9 dairy farms located in the Beijing area of China. Among them, 6,964 cows had MUN records in first and second parities, and 4,991 cows had MUN records in all the first 3 lactations. Milk samples were tested monthly at the Beijing Dairy Cattle Center (Beijing, China), and the pedigree data sets were also provided by the Beijing Dairy Cattle Center. The MUN content was predicted from milk spectra data via internal inbuilt modules in FTS machines (Bentley Instruments Inc.; Lou et al., 2022). Only records within the range of the mean ± 3 standard deviations (SD) and groups with more than 100 records were kept in the data set. The original phenotypic mean for MUN ranged from 1.3 to 25.5 mg/dL (post-quality control ranged from 2.07 to 24.57 mg/dL). The detailed descriptive statistics for the phenotypic records are shown in Table 1. Somatic cell count was obtained through a Fossomatic 5000 machine (Foss Electric) and analyzed as log-transformed SCS.

The pedigree file contained 653,528 animals spanning over 6 generations. A total of 2,029 cows were genotyped using the Illumina 150K Bovine Bead Chip (Illumina Inc.), which contains 121,188 SNP. Samples with low call rates (<0.90) were removed from the data sets. Genotype quality control removed SNP with minor allele frequency lower than 0.05 and in extreme
departure from Hardy-Weinberg equilibrium (P-value lower than 10^{-6}). All genotyped animals were included in genotype imputation analyses for imputing missing SNP using BEAGLE 5.1 software (Browning et al., 2018). Ultimately, 110,009 SNP and 2,029 animals were kept for further analyses.

**Estimation of Genetic Parameters and ssGWAS**

Variance and covariance components and EBV for MUN were calculated using a single-trait repeatability model. The model used can be described as follows:

\[ y = Xb + Z_a a + Z_{pe} e + e, \]

where \( y \) is a vector of phenotypic records; \( b \) is a vector of fixed effects including herd by testing year and month (440 levels), parity (3 levels: 1, 2, and 3–10), and stage of lactation (7 categories based on DIM: 5–50, 51–100, 101–150, 151–200, 201–250, 251–300, or >300); \( X \) is a design matrix associating \( y \); \( a \) is the vector of additive genetic effects; \( e \) is the vector of random residual effects; and \( Z_a \) and \( Z_{pe} \) are the corresponding incidence matrices. The model assumptions are as follows: \( a \sim N(0, \sigma^2_a) \), \( e \sim N(0, \sigma^2_e) \), \( H \) is the relationship matrix combining pedigree and genotype information (Legarra et al., 2009); \( I \) represents an identity matrix; and \( \sigma^2_a, \sigma^2_e, \sigma^2_pe \) are the additive genetic, permanent environment, and residual variances, respectively. The inverse of the \( H \) matrix \((H^{-1})\) was calculated as follows (Aguilar et al., 2010):

\[ H^{-1} = A^{-1} + \begin{bmatrix} \sigma^2_a & 0 \\ \sigma^2_e & \sigma^2_pe \end{bmatrix}, \]

where \( A^{-1} \) is the inverse of the pedigree-based relationship matrix; \( A^{-1}_{aa} \) is the \( A^{-1} \) for the genotyped animals; and \( G^{-1} \) is the inverse of the genomic relationship matrix. The \( G \) matrix was calculated according to VanRaden (2008):

\[ G = \frac{ZDZ'}{2\Sigma P_i (1 - P_i)}. \]

where \( Z \) is the matrix of gene content adjusted for allele frequencies (0, 1, or 2 for aa, Aa, and AA, respectively); \( D \) is a diagonal matrix of weights for SNP variances; \( D \) is a diagonal matrix that corresponds to an identity matrix (I) when \( G \) is fitted considering the same weight (i.e., 1) for all the SNP.

The estimates of effects and P-value for SNP by ssGWAS were obtained according to Aguilar et al. (2019) and Wang et al. (2012). The Manhattan plots showing the significant SNP were created using the “ggplot2” R package (Wilkinson, 2011). To account for multiple testing, 5% chromosome-wise false discovery rate was used to identify significant markers. The inflation factor \( \lambda \) (Devlin and Roeder, 1999) and quantile-quantile (Q-Q) plots were calculated to compare observed distributions of \(-\log(P\text{-value})\) to the expected distribution under the no association model for MUN.

The genetic correlations between MUN and MY, SCS, net energy concentration of the milk (NEm), protein percentage (PP), fat percentage (FP), and lactose percentage (LP) were estimated based on bi-variate repeatability animal models, using the average information (AI)-REML procedure implemented in the DMU package (Madsen et al., 2014). The NEm of each milk sample (NEm in MJ/kg) was calculated as NEm = 0.384 (FP) + 0.223 (PP) + 0.199 (LP) − 0.108 (Tyrrell and Reid, 1965).

The model used for the estimation of (co)variance components and genetic EBV can be described as follows:

\[ y = X\beta + Z_a a + Z_{pe} e + e, \]

where \( y \) is the vector of phenotypic records for MY, SCS, NEm, PP, FP, LP, and MUN; \( \beta \) is the vector of systematic effects, including the same fixed effects described above; \( a \) is the vector of random animal additive genetic effects; \( e \) is the vector of random permanent environmental effects; and \( e \) is the vector

**Table 1.** Descriptive statistics of phenotypic records used in the analyses\(^1\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>NR</th>
<th>NC</th>
<th>Mean</th>
<th>SD</th>
<th>CV%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUN1</td>
<td>94,411</td>
<td>11,503</td>
<td>13.62</td>
<td>3.76</td>
<td>27.6</td>
</tr>
<tr>
<td>MUN2</td>
<td>100,337</td>
<td>11,846</td>
<td>13.60</td>
<td>3.72</td>
<td>27.3</td>
</tr>
<tr>
<td>MUN3</td>
<td>115,505</td>
<td>9,403</td>
<td>13.26</td>
<td>3.75</td>
<td>28.3</td>
</tr>
<tr>
<td>MUN (mg/dL)</td>
<td>399,897</td>
<td>24,435</td>
<td>13.32</td>
<td>3.75</td>
<td>28.2</td>
</tr>
</tbody>
</table>
| MUN1 = milk urea nitrogen in parity 1; MUN2 = milk urea nitrogen in parity 2; MUN3 = milk urea nitrogen in parity 3.

\(^1\text{NR} = \text{number of records}; \text{NC} = \text{number of cows}; \text{MUN1} = \text{milk urea nitrogen in parity 1}; \text{MUN2} = \text{milk urea nitrogen in parity 2}; \text{MUN3} = \text{milk urea nitrogen in parity 3}.\)
of random residual effects; $X$, $Z_a$, and $Z_{pe}$ are the corresponding incidence matrices. The model assumptions are as follows:

$$a \sim N \left(0, A \otimes \begin{bmatrix} \sigma^2_{a(Trait1)} & \sigma_{a(Trait1,Trait2)} \\ \sigma_{a(Trait1,Trait2)} & \sigma^2_{a(Trait2)} \end{bmatrix} \right),$$

$$pe \sim N \left(0, I \otimes \begin{bmatrix} \sigma^2_{pe(Trait1)} & \sigma_{pe(Trait1,Trait2)} \\ \sigma_{pe(Trait1,Trait2)} & \sigma^2_{pe(Trait2)} \end{bmatrix} \right),$$

$$e \sim N \left(0, I \otimes \begin{bmatrix} \sigma^2_{e(Trait1)} & \sigma_{e(Trait1,Trait2)} \\ \sigma_{e(Trait1,Trait2)} & \sigma^2_{e(Trait2)} \end{bmatrix} \right),$$

where $\sigma^2_{k(i,j)}$ is the variance of the effect $k$ ($a$, $pe$, $e$) of trait $i$, and $\sigma_{k(i,j)}$ is the covariance between trait $i$ and $j$ for the effect $k$. $A$ is a pedigree-based relationship matrix and $I$ is an identity matrix ($pe$ is associated with the number of cows and $e$ with the number of records). The heritability and repeatability parameters were defined as $h^2 = \frac{\sigma^2_a}{\sigma^2_p}$ and $r_e = \frac{\sigma^2_e}{\sigma^2_p}$, respectively, where $\sigma^2_p = \sigma^2_a + \sigma^2_{pe} + \sigma^2_e$. PE is the permanent environmental effect when estimating variance components using the single-trait repeated power model; $pe$ is the permanent environmental effect when estimating genetic correlations between MUN and MY, SCS, NE, PP, FP, and LP using the multitrait repeated power model. The single-trait repetitive force model used is blupf90; the multitrait repetitive force model used is DMU. The squares of the standard errors (SE) for the heritability and genetic correlation coefficient were calculated as follows (Su et al., 2007):

$$SE^2(h^2) = \left[ \frac{\sigma^2_a}{\sigma^2_p} \right]^2 \left[ \frac{\text{Var}(\hat{\sigma}^2_a)}{\hat{\sigma}^2_a} + \frac{\text{Var}(\hat{\sigma}^2_p)}{\hat{\sigma}^2_p} - 2 \frac{\text{Cov}(\hat{\sigma}^2_a, \hat{\sigma}^2_p)}{\sigma^2_a \sigma^2_p} \right].$$

The genetic correlation coefficient was calculated as $r = \frac{\hat{\sigma}_{ij}}{\sqrt{\sigma^2_{ij} \sigma^2_{ij}}}$, and the square of the SE for the genetic correlation coefficient was calculated as follows (Su et al., 2007):

$$SE^2(r) = \left[ \frac{\sigma^2_{ij}}{\sigma^2_{ij} \sigma^2_{ij}} \right] \left[ \frac{\text{Var}(\hat{\sigma}_{ij}) + \text{Var}(\hat{\sigma}^2_i) + \text{Var}(\hat{\sigma}^2_j)}{\hat{\sigma}_{ij}^2} - 4(\hat{\sigma}^2_{ij})^2 \frac{\hat{\sigma}_{ij}^2}{4(\hat{\sigma}^2_i)^2} + \frac{\text{Cov}(\hat{\sigma}_{ij}, \hat{\sigma}^2_i) + \text{Cov}(\hat{\sigma}_{ij}, \hat{\sigma}^2_j)}{\sigma^2_{ij} + 2\sigma^2_{ij} \sigma^2_{ij}} \right].$$

The variances and covariances of the estimated parameters $(\sigma^2_i, \sigma^2_j)$ were obtained from the inverse of the average information matrix. The reliability of the EBV was calculated as

$$\text{Reliability} = 1 - \frac{\text{SEP}^2}{\sigma^2_a},$$

where SEP is the standard error of prediction and $\sigma^2_a$ is the direct additive genetic variance. The data sets used to estimate genetic parameters and EBV for the traits described herein were extracted in April 2021.

Identification of Candidate Genes, Quantitative Trait Loci, and Functional Annotation

The SNP chip panel used had an average intermarker spacing of 26.07 kb, and the average distance between adjacent SNP with linkage disequilibrium (measured as $r^2$) higher than 0.4 was 200 kb in the studied population. Therefore, the significant SNP were surveyed to their corresponding genes or to surrounding genes within a distance of 200 kb upstream or downstream from the significant SNP. Positional genes were annotated based on the start and end coordinates of each genomic window (considering the ARS-UCD1.2 assembly as the reference genome) using the Ensembl R package “Biomart” (Durinck et al., 2009). We scanned the corresponding QTL regions within the 200-kb genomic region surrounding the significant SNP based on the Animal QTL database (www.animalgenome.org/cgi-bin/QTLdb/index; Hu et al., 2022, 2007). To better understand the biological processes and pathways shared by these candidate genes, we also conducted enrichment analyses. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) terms were enriched via the “clusterProfiler” package (Yu et al., 2012).

RESULTS

Descriptive Statistics

The phenotypic means (SD) of MUN were 13.62 (3.76), 13.60 (3.72), and 13.26 (3.75) mg/dL for the
first, second, and third parity, respectively. The trends of MUN with DIM increasing for different parities were similar (Figure 1). Levels of MUN gradually increased during early lactation (0–100 DIM) and reached a maximum within 100 to 150 DIM. During the late stage of lactation, MUN for first- and second-parity cows showed a downward trend. However, we found an upward trend for third-parity cows, which may be due to the smaller number of MUN records after 300 d for cows with 3 or more parities. Descriptive statistics of MUN concentration for the first 3 parities are shown in Supplemental Table S1 (https://doi.org/10.6084/m9.figshare.20510328.v2; Ma and Wang, 2022).

**Genetic Parameters and Genetic Correlation with Milk Performance**

The heritability estimates of MUN were 0.06 ± 0.004 (additive genetic variance = 0.47; permanent environmental variance = 0.54; residual variance = 7.31) with a repeatability of 0.12. The phenotypic and genetic correlations for MUN with MY, SCS, NEm, FP, PP, and LP are presented in Table 2. As additional information, we also provide descriptive statistics for the data sets used to estimate genetic correlations in Supplemental Table S1 and the genetic parameter estimates using a 2-trait repeatability model in Supplemental Table S2 (https://doi.org/10.6084/m9.figshare.20510328.v2; Ma and Wang, 2022). Based on the bivariate repeatability animal models, the phenotypic correlations of MUN with MY, NEm, FP, and LP were positive and ranged from 0.03 to 0.08, and negative correlations with PP and SCS were −0.03 and −0.10, respectively. The genetic correlations between MUN and MY, NEm, FP, PP, and LP were positive and ranged from 0.02 to 0.26. We found a negative genetic correlation of −0.18 between MUN and SCS.

**ssGWAS and Functional Enrichment Analyses**

The Manhattan and Q-Q plots for the ssGWAS results of MUN are shown in Figure 2 and Supplemental Figure S1 (https://doi.org/10.6084/m9.figshare.20510328.v2; Ma and Wang, 2022). A total of 18 genome-wide significant SNP (Table 3), located on BTA11, BTA12, BTA14, BTA17, and BTA18, were identified for MUN. The minor allele frequency of the significant SNP ranged from 0.17 (ARS-BFGL-NGS-118057) to 0.48 (BovineHD1100030069), and the adjusted P-value ranged from 2.59 × 10−6 to 0.02 (BovineHD1100029855). The Q-Q plot and the lambda value indicate that potential population stratification seems to have been accounted for in the GWAS analyses. It was expected that a larger number of significant SNP would be identified. However, the results observed might be due to the highly polygenic nature of nitrogen

### Table 2. Genetic and phenotypic correlations between MUN and milk yield (MY), net energy concentration of milk (NEm; Tyrrell and Reid, 1965), log-transformed SCS, fat percentage (FP), protein percentage (PP), and lactose percentage (LP) in Holstein cattle

<table>
<thead>
<tr>
<th>Trait</th>
<th>Phenotypic correlation</th>
<th>Genetic correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MY</td>
<td>0.05 ± 0.05</td>
<td>0.07 ± 0.06</td>
</tr>
<tr>
<td>NEm</td>
<td>0.04 ± 0.03</td>
<td>0.15 ± 0.05</td>
</tr>
<tr>
<td>SCS</td>
<td>−0.10 ± 0.11</td>
<td>−0.18 ± 0.05</td>
</tr>
<tr>
<td>FP</td>
<td>0.03 ± 0.03</td>
<td>0.26 ± 0.05</td>
</tr>
<tr>
<td>PP</td>
<td>−0.03 ± 0.04</td>
<td>0.02 ± 0.05</td>
</tr>
<tr>
<td>LP</td>
<td>0.08 ± 0.08</td>
<td>0.14 ± 0.05</td>
</tr>
</tbody>
</table>

1Values shown are correlation coefficients ± SE.
conversion capacity; each variant has a small effect (and are more difficult to identify) in the phenotype evaluated, and the heritability estimated was MUN. By searching for flanking regions ±200 kb on either side of significant SNP, 6 candidate genes, including CFAP77, CAMSAP1, CACNA1B, FARPI, ADGRB1, and INTU, covering or adjacent to 18 SNP, were detected in BTA11, BTA12, BTA14, and BTA17 associated with MUN. A total of 1,797 previously reported QTL were identified in the significant genomic regions, most of which are located on BTA11. Few QTL were located in BTA12, BTA14, BTA17, and BTA18. Surprisingly, except for the QTL for interval to first estrus after calving, the 16 identified QTL regions were all associated with milk quality, BCS, and milk protein. The details of the mapped QTL regions are presented in Table 3.

Based on the association results, 18 significant SNP were used to assign 67 unique genes within 200 kb flanking regions of these SNP. Supplemental Table S3 (https://doi.org/10.6084/m9.figshare.20510328.v2; Ma and Wang, 2022) presents the KEGG and GO results. Chordate embryonic development (GO:0043009), embryo development ending in birth or egg hatching (GO: 0009792), epithelial cell differentiation (GO:0030855), organelle assembly (GO:0070925), and embryo development (GO:0009790) were the enriched biological process with the highest number of genes from the input data set. Meanwhile, GO:0043009 (chordate embryonic development) was the most significantly enriched biological process for MUN positional candidate genes ($P = 0.02$). In addition, 6 KEGG pathways were also enriched for MUN candidate genes: cortisol synthesis and secretion, steroid hormone biosynthesis, glycosphingolipid biosynthesis–globo and isoglobo series, thiamine metabolism, steroid biosynthesis, and Cushing syndrome. Cortisol synthesis and secretion was the most significantly enriched pathway ($P = 0.0094$).

DISCUSSION

Descriptive Statistics of MUN

Intensive genetic selection for productive traits over the past 50 years has led to adjustment of cows’ metabolism to distribute more nutrients and energy toward milk synthesis (Miglior et al., 2017). Milk urea nitrogen is an important indicator for evaluating the utilization of protein in the dairy cow’s diet and the balance of protein and energy in the diet. Meanwhile, lower MUN could also be linked with improved N use efficiency. However, under the hypothesis that reduced MUN has no effect on the amount of protein produced (Jonker et al., 1998), some researchers have demonstrated the direct interest in breeding for lower MUN to reduce the
environmental impact of dairy production (Bobbo et al., 2020; Marshall et al., 2020).

The average MUN found in this study (13.62, 13.60, and 13.26 mg/dL in the first, second, and third lactation, respectively) is lower than those reported in previous studies (Miglior et al., 2006; Bastin et al., 2009; Mucha and Strandberg, 2011) and lower than the range of 15.0 to 30.0 mg/dL proposed by Glatz-Hoppe et al. (2020).

Chen et al. (2021a) reported that the mean MUN in the first 3 lactations of Walloon Holstein cows ranged from 24.80 to 26.19 mg/dL, which is slightly higher than our results in Chinese Holstein cattle. The mean MUN found in the first parity was slightly higher than that found in the second parity, which is consistent with previous studies (de Roos and de Jong, 2006; Cao et al., 2010; Rzewuska and Strabel, 2013; Atashi and Hostens, 2021). Compared with primiparous cows, lower MUN concentration in multiparous cows may be associated with higher BW and greater MY level (Jonker et al., 1998). The coefficient of variations (CV) for MUN were 27.6, 27.3, and 28.3% in the first, second, and third parity, respectively. The CV for MUN were 27% in Italian Brown Swiss dairy cows (Samoré et al., 2007), 37% in grazing dairy cattle in New Zealand (Lopez-Villalobos et al., 2018), 26% in Belgian Holsteins (Atashi and Hostens, 2021), 33% in Dutch Holsteins (Stoop et al., 2007), and ranged from 40 to 41% in Polish Holstein cows (Samoré et al., 2007).

### Heritability and Genetic Correlation

Milk urea nitrogen can be measured at a low cost. Based on our results, we find enough additive genetic variance to enable genetic progress for MUN. The ssGWAS results obtained show that MUN is a polygenically controlled trait. Yazgan et al. (2010) reported test-day heritabilities for MUN ranging from 0.14 to 0.25 in Polish Holstein cows. Mitchell et al. (2005) estimated heritability of 0.22 based on 26,540 first-parity Holstein cows records using infrared spectroscopy, and 0.14 using wet chemistry techniques to determine MUN in 25,902 first-parity Holstein cows. The estimates of heritability for MUN in our study are consistent with those reported by Mitchell et al. (2005) and Mitchell et al. (2018), with heritabilities of 0.22 for the first and second parity, respectively, and 0.24 for the first and second parity, respectively. We speculate that these differences in the heritability estimates of MUN might reflect the genetic variability across breeds, as well as differences in data collection protocols, trait definition, and statistical models used. Stoop et al. (2007) reported a MUN heritability of 0.14 based on data from 1,953 primiparous Holstein cows. Wenninger and Distl (1993) estimated heritabilities for milk urea in German Brown Swiss dairy cows (Samoré et al., 2007).

### Table 3. Significant SNP and candidate genes identified for MUN in lactating Holstein cows

<table>
<thead>
<tr>
<th>SNP</th>
<th>BTA</th>
<th>Position (bp)</th>
<th>P-value</th>
<th>FDR</th>
<th>Candidate genes (within ± 200 kb)</th>
<th>Main QTL types (number)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BovineHD11000027 55</td>
<td>11</td>
<td>102,536,578</td>
<td>2.31 × 10^{-6}</td>
<td>0.018157</td>
<td>CFAP77</td>
<td>Milk β-LG protein content QTL (77)</td>
</tr>
<tr>
<td>BovineHD11000027 97</td>
<td>11</td>
<td>102,603,890</td>
<td>2.18 × 10^{-6}</td>
<td>0.002992</td>
<td>CFAP77</td>
<td>Milk β-LG protein content QTL (79)</td>
</tr>
<tr>
<td>BovineHD11000027 97</td>
<td>11</td>
<td>102,657,358</td>
<td>7.58 × 10^{-6}</td>
<td>0.000208</td>
<td>CFAP77</td>
<td>Milk β-LG protein content QTL (85)</td>
</tr>
<tr>
<td>ARS-BFG-N1-31007</td>
<td>11</td>
<td>103,289,035</td>
<td>3.82 × 10^{-6}</td>
<td>0.001675</td>
<td>CFAP77</td>
<td>Milk β-LG protein content QTL (68)</td>
</tr>
<tr>
<td>BovineHD11000030 09</td>
<td>11</td>
<td>103,304,643</td>
<td>4.96 × 10^{-6}</td>
<td>0.002072</td>
<td>CFAP77</td>
<td>Milk β-LG protein content QTL (157)</td>
</tr>
<tr>
<td>BovineHD11000030 78</td>
<td>11</td>
<td>103,326,092</td>
<td>9.79 × 10^{-6}</td>
<td>0.000406</td>
<td>CFAP77</td>
<td>Milk β-LG protein content QTL (157)</td>
</tr>
<tr>
<td>BovineHD11000030 13</td>
<td>11</td>
<td>103,456,996</td>
<td>1.52 × 10^{-6}</td>
<td>0.002391</td>
<td>CAMSAP1</td>
<td>Milk β-LG protein content QTL (147)</td>
</tr>
<tr>
<td>ARS-BFG-N1-106343</td>
<td>11</td>
<td>106,804,258</td>
<td>1.81 × 10^{-6}</td>
<td>0.019172</td>
<td>CAMSAP1</td>
<td>Milk β-LG percentage QTL (9)</td>
</tr>
<tr>
<td>BovineHD1200021008</td>
<td>12</td>
<td>75,346,943</td>
<td>2.09 × 10^{-6}</td>
<td>0.01772</td>
<td>FAR1</td>
<td>Milk fat percentage QTL (912)</td>
</tr>
<tr>
<td>BovineHD1200021827</td>
<td>12</td>
<td>77,231,205</td>
<td>1.97 × 10^{-6}</td>
<td>0.019172</td>
<td>ADGRB1</td>
<td>Milk butyric acid content QTL (1)</td>
</tr>
<tr>
<td>ARS-BFG-N1-4939</td>
<td>14</td>
<td>1,801,116</td>
<td>6.93 × 10^{-6}</td>
<td>0.045399</td>
<td>INTU</td>
<td>Milk butyric acid content QTL (2)</td>
</tr>
<tr>
<td>ARS-USMARC-Parent-AYS55890-rs2900256</td>
<td>17</td>
<td>29,926,334</td>
<td>2.36 × 10^{-6}</td>
<td>2.5E-05</td>
<td>INTU</td>
<td>Milk β-LG protein content QTL (10)</td>
</tr>
<tr>
<td>BovineHD11000028 55</td>
<td>11</td>
<td>30,063,422</td>
<td>4.90 × 10^{-6}</td>
<td>0.001079</td>
<td>INTU</td>
<td>Milk β-LG percentage QTL (9)</td>
</tr>
<tr>
<td>BovineHD1700008492</td>
<td>17</td>
<td>30,098,730</td>
<td>1.41 × 10^{-6}</td>
<td>0.002391</td>
<td>INTU</td>
<td>Milk fat percentage QTL (912)</td>
</tr>
<tr>
<td>BovineHD1700008502</td>
<td>17</td>
<td>30,135,486</td>
<td>7.43 × 10^{-6}</td>
<td>0.045399</td>
<td>INTU</td>
<td>Milk butyric acid content QTL (2)</td>
</tr>
<tr>
<td>ARS-USMARC-547</td>
<td>17</td>
<td>30,401,823</td>
<td>4.06 × 10^{-6}</td>
<td>0.029774</td>
<td>INTU</td>
<td>Milk β-LG protein content QTL (10)</td>
</tr>
<tr>
<td>BovineHD11000029 55</td>
<td>11</td>
<td>30,623,157</td>
<td>7.78 × 10^{-6}</td>
<td>4.2E-05</td>
<td>INTU</td>
<td>Milk β-LG percentage QTL (9)</td>
</tr>
<tr>
<td>ARS-BFG-N1-40805</td>
<td>17</td>
<td>30,663,422</td>
<td>4.90 × 10^{-6}</td>
<td>0.001079</td>
<td>INTU</td>
<td>Milk fat percentage QTL (912)</td>
</tr>
<tr>
<td>BovineHD1700008492</td>
<td>17</td>
<td>30,098,730</td>
<td>1.41 × 10^{-6}</td>
<td>0.002391</td>
<td>INTU</td>
<td>Milk fat percentage QTL (912)</td>
</tr>
<tr>
<td>ARS-USMARC-547</td>
<td>17</td>
<td>30,401,823</td>
<td>4.06 × 10^{-6}</td>
<td>0.029774</td>
<td>INTU</td>
<td>Milk β-LG protein content QTL (10)</td>
</tr>
<tr>
<td>BovineHD1800004197</td>
<td>18</td>
<td>13,623,153</td>
<td>7.02 × 10^{-6}</td>
<td>0.045399</td>
<td>INTU</td>
<td>Milk β-LG percentage QTL (9)</td>
</tr>
</tbody>
</table>

1Positions of the SNP and genes were based on the Ensembl Genes 104, Cow genes (ARS-UCD1.2) database (www.ensembl.org/index.html). FDR = false discovery rate.
Simmental cows of 0.06 ± 0.04 and in German Brown cows of 0.25 ± 0.11. Wood et al. (2003) estimated a higher heritability of 0.44 for MUN, using a random regression model with at least 4 test-day samples per cow and fitting heterogeneous residual variance for 6,704 first-parity Holstein cows. The MUN heritability in German Simmentals is similar to that of Chinese Holsteins. Chen et al. (2021a,b) used MUN data of 560,739 Holstein cows and a random regression model to estimate the MUN heritability in 3 parities, which were 0.19 ± 0.02 for the first parity, 0.22 ± 0.02 for the second parity, and 0.20 ± 0.02 for the third parity. The differences in the estimates obtained may be due to the sample size and statistical models used.

The phenotypic relationships of MUN with MY and milk composition traits were weak, which is in agreement with the literature (Samoré et al., 2007; Stoop et al., 2007; Miglior et al., 2017). Phenotypic correlations of MUN with MY, FP, PP, LP, and SCS ranged from -0.08 to 0.07. The estimated genetic correlations between MUN and NEm, MY, PP, LP, and MUN were positive and ranged from 0.02 ± 0.05 to 0.26 ± 0.05, except for a negative genetic correlation between MUN and SCS (-0.18 ± 0.05). Our findings are in agreement with those of Samoré et al. (2007), who reported a positive (0.12) genetic correlation between FP and MUN in Brown Swiss cattle. Miglior et al. (2017) reported a weak negative genetic correlation between MUN and MY (-0.09), whereas moderate positive genetic correlations were estimated between MUN and FP (0.42) and MUN and PP (0.20) in Canadian Holstein cattle. Wood et al. (2003) reported genetic correlations of MUN with MY, PP, and FP that were also generally weak (-0.05 to 0.17), except for those of second-parity MUN with FP (0.32) and PP (0.22). However, other authors (Stoop et al., 2007; Rzewuska and Strabel, 2013; Satola et al., 2017) have reported null or weak genetic correlation between MUN and FP. It is noteworthy that the SE (range from 0.10 to 0.16) of the 2- and 3-parity estimates reported by Wood et al. (2003) are larger than those of the present study (range from 0.05 to 0.06), except for the SE of first-parity MUN (0.04).

Weak negative genetic correlations of -0.19 and -0.08 between MUN and SCS were reported by Miglior et al. (2017) and Samoré et al. (2007), respectively, which is consistent with our results. Mucha and Strandberg (2011) reported positive genetic correlations at the beginning and negative genetic correlations at the end of the lactation between MUN and MY-related traits (MY, FY, and PY). König et al. (2008) reported a moderate positive genetic correlation between MU and MY. Because NEm reflects the energy concentration of the milk, this trait might be related to the energy status of the cow, as is MUN (Jorritsma et al., 2003). However, the genetic correlation between MUN and NEm was low (0.15 ± 0.05). Studies have reported that selection for increasing milk production traits and decreasing SCS could lead to decreasing MUN. The approximate genetic correlation between MU and SCS ranged from -0.20 to 0.28. The range of approximate genetic correlations between MU and production traits (MY, FY, and PY) was found to be from -0.25 to -0.01. Chen et al. (2021a,b) have reported approximate genetic correlations of MUN with FP and PP ranging from 0.14 to 0.26 with SE of 0.03, which are similar to our results (range from 0.02 ± 0.05 to 0.26 ± 0.05). We hypothesize that the differences in the results of PP may be due to breed differences, as similar SE estimates indicate that our results are similarly reliable. The heritability of MY estimated based on a bivariate repeatability model was lower than expected, which reflects greater environmental variability in the farms involved in the study.

Previous research has shown that MUN can be used as an indicator of urinary urea nitrogen (UUN) excretion level of dairy cows in late lactation (Barros et al., 2019) and, therefore, a potential trait for reducing environmental footprints of dairy production. A positive relationship (16.01% and 18.67%) has been observed between MUN and UUN (Jonker et al., 1998; Burgos et al., 2007). This indicates that genetic selection for reducing MUN might result in an indirect response in UUN. On the premise of not affecting productive performance, breeding for reduced MUN will not only help to reduce the cost of dairy cattle production but also minimize the environmental footprint of the dairy cattle industry.

**Trait-Related Genes**

Six positional genes related to protein metabolism or binding to specific proteins, including **CFAP77**, **CAMSAP1**, **CACNA1B**, **FARPI**, **ADGRB1**, and **INTU**, were suggested to be candidate genes associated with MUN in Holstein cattle. Our study is limited by the reduced number of genotyped individuals. Therefore, increasing the size of the reference population and determining the concentration of each protein in milk could effectively improve the ability to detect MUN-related genomic regions. The identified genomic regions contribute to the understanding of biological mechanisms influencing MUN and can also be used when designing new SNP chip panels (including additional SNP in the identified regions) to improve the accuracy of genomic predictions of breeding values for MUN.

Bouwman et al. (2010) reported a heritability estimate for MUN of 0.14, which is higher than the one observed in the Chinese Holstein population. Furthermore, the QTL regions identified by those authors were
located on BTA1, BTA6, BTA21, and BTA23, which do not overlap with the regions identified in the current study. This might be due to the highly polygenic nature of MUN and different methods used. This is in line with the study by Ariyarathne et al. (2021), which performed GWAS for MUN in dairy cows (Holstein, Friesian, Jersey, and crossbred cows) raised in grazing systems and did not identify any significant SNP. van den Berg et al. (2022) reported SNP significantly associated with MUN, which are located on BTA11 (position = 103,271,858 and 105,520,434; \( P \)-value in Australian Holsteins = \( 5.4 \times 10^{-6} \) and \( 3.8 \times 10^{-4} \); \( P \)-value in New Zealand Holsteins = \( 8.8 \times 10^{-7} \) and \( 7.4 \times 10^{-5} \)). When searching for genes 1 Mb upstream and downstream of the significant SNP located on BTA11 in this study, we also mapped the same significant genes GLT6D1 and ENSBTAG00000054425-OLFM1 reported in their study (van den Berg et al., 2022). There are indeed important SNP that have an effect on MUN but do not reach a significant level, van den Berg et al. (2022) found some patterns in QTLD detected in the meta-analysis for MUN. The 2 SNP found on BTA11 were located very close to the 8 significant SNP on BTA11 in our study. Not all the same genes were identified, because we only searched for genes in the vicinity of 200 kb of the significant SNP, whereas van den Berg et al. (2022) considered a window of 1 Mb. We identified QTLD regions associated with milk production traits on BTA11 and BTA17, which have not been previously reported in other studies. These QTLD regions are milk \( \beta \)-LG protein content QTLD, milk \( \beta \)-LG protein content QTLD, milk \( \beta \)-LG percentage QTLD, milk FP QTLD, and milk butyric acid content QTLD. These regions are all related to milk protein, indicating the importance of further exploring the relationship between MUN and PP.

Two significant SNP were mapped to the same candidate gene (CFAP77, cilia and flagella associated protein 77) located on BTA11, which is a protein coding gene. The CFAP77 gene has been predicted to be located in cilia of water buffalo (Du C et al., 2019). Montalvo-Ortiz et al. (2019) reported that CFAP77 is involved in chromatin remodeling, DNA binding, cell survival, and cell projection. Also, on BTA11, the CAMSAP1 (calmodulin regulated spectrin associated protein 1) gene contains a significantly associated SNP in its intronic region. Gene Ontology annotations related to this gene include microtubule binding and spectrin binding. King et al. (2014) indicated that CAMSAP1 is a vertebrate microtubule-binding protein that plays an important role in animal growth and development and protein metabolism.

On BTA11, CACNA1B (calcium voltage-gated channel subunit alpha1 B) is a protein coding gene involved in a variety of calcium-dependent processes, including muscle contraction, hormone or neurotransmitter release, gene expression, cell motility, cell division, and cell death. Hu et al. (2021) reported that CACNA1B is involved in the glycosaminoglycan biosynthesis pathway. Interestingly, MUN was genetically correlated with LP and PP in our studied population, indicating that CACNA1B might be involved in MUN production.

On BTA12, FARPI encodes a protein containing a FERM domain, a Dbh homology domain, and 2 pleckstrin homology domains. These domains are found in guanine nucleotide exchange factors and proteins that link the cytoskeleton to the cell membrane. FARPI (FERM, ARH/RhoGEF, and pleckstrin domain protein 1) is a protein coding gene, which encodes an important accessory protein for protein metabolism (Cheadle and Biederer, 2012). Milk urea nitrogen is an important intermediate product in the conversion of crude protein to nonprotein nitrogen, so we infer that FARPI may be a regulator of MUN production or metabolism. At the same time, FARPI was also enriched for the chordate embryonic development pathway, indicating that MUN might be a potential reference indicator for measuring chordate embryonic development that needs to be further investigated.

The ADGRB1 gene (adhesion G protein-coupled receptor B1) located on BTA14 is a protein coding gene. Zhu et al. (2018) reported that 2 binding sites of ADGRB1 play a decisive role in angiogenesis. Angiogenesis is controlled by a local balance between stimulators and inhibitors of new vessel growth and is suppressed under normal physiologic conditions. Blood UN is formed by the filtration of mammary glands. Therefore, ADGRB1, which affects angiogenesis, might be related to the generation and transportation of blood UN. In this context, ADGRB1 might have an indirect effect on MUN.

On BTA17, 2 significant SNP were mapped to an important gene (INTU; inturned planar cell polarity protein), which is a protein coding gene involved in ciliogenesis and embryonic development. INTU regulates cilia formation by controlling the organization of the apical actin cytoskeleton and the positioning of the basal bodies at the apical cell surface, which in turn is essential for the normal orientation of elongating ciliary microtubules. In a study of human tissue function, ablation of INTU specifically in kidney proximal tubules was shown to aggravate renal ischemia-reperfusion injury, leading to defective post-injury ciliogenesis and renal metabolic disorder (Wang et al., 2018). Meanwhile, it has been reported that MUN is a predictor of UN excretion (Spek et al., 2013). The INTU gene was also enriched in the chordate embryonic development pathway. Based on the functions of the FRP1 and INTU genes and the relationship between MUN and UN, we believe
that \textit{FRP1} and \textit{INTU} are regulatory genes involved in the production and metabolism of MUN. The 2 genes jointly regulate the level of MUN in the body, thereby reflecting the growth and development status (nutrient level) of the body and the level of UN excretion. Some other enriched pathways related to growth and development seem to be related to similar processes (e.g., embryo development ending in birth or egg hatching, epithelial cell differentiation, organelle assembly, and embryo development). Cortisol synthesis and secretion, steroid hormone biosynthesis, glycosphingolipid biosynthesis - globo and isoglobo series, and other significant pathways were enriched. These pathways are related to hormone metabolism, cortisol, and steroid hormones, which are generally important for blood metabolism regulation. Because the source of MUN is BUN, we speculate that these pathways can affect the generation of MUN. Overall, these genes significantly associated with MUN are involved in protein catabolism, the urea cycle, ion transportation, and N excretion.

Six positional genes related to protein metabolism or binding to specific proteins were suggested to be candidate genes associated with MUN in Holstein cattle. We recommend establishing a unified standard for the selection index of N conversion efficiency in dairy cows, including indicators, measurement methods, and optimal concentration range, which is one of our future goals.

\textbf{Implications and Future Studies}

Our results support MUN as a useful tool for monitoring N emissions in dairy cows. Milk urea nitrogen is heritable with enough additive genetic variation to enable genetic progress for N emissions. It is also important to evaluate additional physiological indicators of N efficiency, which is an area that our group is currently working on. In the long term, we expect to develop a selection subindex for high N conversion capacity based on multiple indicators of N efficiency. The genomic regions identified in this population also need to be validated in other Holstein cattle populations as well as in other breeds. The SNP and genomic regions identified in this study can be used to improve the accuracy of genomic prediction for high N conversion capacity. For instance, the key SNP identified can be included in low-density SNP panels to increase accuracy of genomic predictions based on imputed genotypes.

\textbf{CONCLUSIONS}

Genetic parameters of MUN in Holstein cattle were estimated using a single-trait repeatability model. Milk urea nitrogen is a trait with low heritability and repeatability. The correlations of MUN with milk performance traits were low to moderate, but unfavorable in some cases. These results indicate that genetic progress for reduced MUN can be achieved through direct genetic selection. However, as MUN is genetically related to some production traits, these relationships should not be ignored when designing selection schemes. The numerous genomic regions identified on different chromosomes explaining small proportions of the total additive genetic variance of MUN indicate that this is a highly polygenic trait. Six genes (\textit{CFAP77}, \textit{CAMSAPI}, \textit{CACA1B}, \textit{ADGRB1}, and \textit{INTU}) are the main candidate genes influencing MUN in Holstein cows.

\textbf{ACKNOWLEDGMENTS}

This research was funded by the earmarked fund for CARS-36 and the Program for Changjiang Scholar and Innovation Research Team in University (IRT_15R62; China). The authors are grateful to the Beijing Dairy Cattle Center (BDCC; Beijing, China) and Beijing Sunlon Livestock Development Company Limited (Beijing, China) for facilitating data collection and providing access to production records. We also acknowledge the members of the 459 Laboratory at China Agriculture University (Beijing) for help with data collection. This research was funded by the China Agriculture Research System of MOF and MARA (CARS-36); the Program for Changjiang Scholar, and Innovation Research Team in University (IRT_15R62). Author Gang Guo is employed by Beijing Sunlon Livestock Development Co., Ltd. The author is involved in custody of the data of dairy herd improvement program. The authors have not stated any other conflicts of interest.

\textbf{REFERENCES}


ORCIDs

Longguang Ma https://orcid.org/0000-0002-3813-9476
Hanpeng Luo https://orcid.org/0000-0001-6221-3834
Luiz F. Brito https://orcid.org/0000-0002-5819-0922
Yao Chang https://orcid.org/0000-0002-3096-1027
Ziwei Chen https://orcid.org/0000-0003-4076-8164
Wenqi Lou https://orcid.org/0000-0002-0150-6068
Gang Guo https://orcid.org/0000-0001-8677-6325
Yachun Wang https://orcid.org/0000-0003-3629-2802

Journal of Dairy Science Vol. 106 No. 1, 2023