

1 **Validating genomic prediction for nitrogen efficiency index and its composition**
2 **traits of Holstein cows in early lactation**

3 Y. Chen^{1*}, H. Atashi^{1,2}, R. R. Mota³, C. Grelet⁴, S. Vanderick¹, H. Hu¹, GplusE Consortium[†],
4 and N. Gengler¹

5

6 ¹ TERRA Teaching and Research Center, University of Liège, Gembloux Agro-Bio Tech
7 (ULiège-GxABT), 5030 Gembloux, Belgium

8 ² Department of Animal Science, Shiraz University, 71441-65186 Shiraz, Iran

9 ³ Council on Dairy Cattle Breeding, Bowie, 20716 MD, United States of America

10 ⁴ Walloon Agricultural Research Center (CRA-W), 5030 Gembloux, Belgium

11

12 [†] List of authors within the GplusE consortium: Mark Crowe, Alan Fahey, Fiona Carter, Elizabeth
13 Matthews, Andreia Santoro, Colin Byrne, Pauline Rudd, Roisin O’Flaherty, Sinead Hallinan,
14 Claire Wathes, Mazdak Salavati, Zhangrui Cheng, Ali Fouladi, Geoff Pollott, Dirk Werling,
15 Beatriz Sanz Bernardo, Conrad Ferris, Alistair Wylie, Matt Bell, Mieke Vaneetvelde, Kristof
16 Hermans, Miel Hostens, Geert Opsomer, Sander Moerman, Jenne De Koster, Hannes Bogaert, Jan
17 Vandepitte, Leila Vandevælde, Bonny Vanranst, Klaus Ingvarsten, Martin Tang Sorensen, Johanna
18 Hoglund, Susanne Dahl, Soren Ostergaard, Janne Rothmann, Mogens Krogh, Else Meyer, Leslie
19 Foldager, Charlotte Gaillard, Jehan Ettema, Tine Rousing, Torben Larsen, Victor H. Silva de
20 Oliveira, Cinzia Marchitelli, Federica Signorelli, Francesco Napolitano, Bianca Moioli,
21 Alessandra Crisà, Luca Buttazzoni, Jennifer McClure, Daragh Matthews, Francis Kearney,
22 Andrew Cromie, Matt McClure, Shujun Zhang, Xing Chen, Huanchun Chen, Junlong Zhao, Ligu

Validating genomic prediction for nitrogen efficiency index

23 Yang, Guohua Hua, Chen Tan, Guiqiang Wang, Michel Bonneau, Marlène Sciarretta, Armin
24 Pearn, Arnold Evertson, Linda Kosten, Anders Fogh, Thomas Andersen, Matthew Lucy, Chris
25 Elsik, Gavin Conant, Jerry Taylor, Deborah Triant, Nicolas Gengler, Michel Georges, Frederic
26 Colinet, Marilou Ramos Pamplona, Hedi Hammami, Catherine Bastin, Haruko Takeda, Aurelie
27 Laine, Anne-Sophie Van Laere, Rodrigo Mota, Saeid Naderi Darbagshahi, Frederic Dehareng,
28 Clement Grelet, Amelie Vanlierde, Eric Froidmont, Frank Becker, Martin Schulze, Sergio Palma
29 Vera.

30

31 *Corresponding Author: Yansen Chen
32 Passage des Déportés, 2
33 B-5030 Gembloux
34 Belgium
35 +32/81/62 23 58 (office)
36 yansen.chen@uliege.be (e-mail)

37 **Abstract**

38 Nitrogen (N) use efficiency (NUE) is an economically important trait for dairy cows.
39 Recently, we proposed a new N efficiency index (NEI), that simultaneously considers both
40 NUE and N pollution. This study aimed to validate the genomic prediction for NEI and its
41 composition traits and investigate the relationship between SNP effects estimated directly
42 from NEI and indirectly from its composition traits. The NEI composition included genomic
43 estimated breeding value of N intake (NINT), milk true protein N (MTPN), and milk urea
44 N yield. The edited data were 132,899 records on 52,064 cows distributed in 773 herds. The
45 pedigree contained 122,368 animals. Genotypic data of 566,294 SNP was available for 4,514
46 individuals. A total of 4,413 cows (including 181 genotyped) and 56 bulls (including 32
47 genotyped) were selected as the validation populations. The linear regression method was
48 used to validate the genomic prediction of NEI and its composition traits using best linear
49 unbiased prediction (BLUP) and single-step genomic BLUP (ssGBLUP). The mean
50 theoretical accuracies of validation populations obtained from ssGBLUP were higher than
51 those obtained from BLUP for both NEI and its composition traits, ranging from 0.57
52 (MTPN) to 0.72 (NINT). The highest mean prediction accuracies for NEI and its
53 composition traits were observed for the genotyped cows estimated under ssGBLUP,
54 ranging from 0.48 (MTPN) to 0.66 (NINT). Furthermore, the SNP effects estimated from
55 NEI composition traits, multiplied by the relative weight were the same as those estimated
56 directly from NEI. This study preliminary showed that genomic prediction can be used for
57 NEI, however, we acknowledge the need for further validation of this result in a larger
58 dataset. Moreover, the SNP effects of NEI can be indirectly calculated by using the SNP

59 effects estimated from its composition traits. This study provided a basis for adding genomic
60 information to establish NEI as part of future routine genomic evaluation programs.

61 **Keywords:** nitrogen use efficiency, SNP effect, mid-infrared spectra

62

Introduction

63 Livestock production contributes to 18% of the global human-induced greenhouse gas
64 emissions in the form of carbon dioxide, methane, and nitrous oxide (Moran and Wall, 2011).
65 The main sources of nitrous oxide production in cattle are deposited urine and manure (Rochette
66 et al., 2014), primarily due to excessive protein content in their feed. Protein constitutes the
67 costliest component of indoor dairy cow feed, and protein loss reduces farm profitability. Thus,
68 improving the nitrogen (**N**) use efficiency (**NUE**) of dairy cows has the potential to enhance the
69 profitability of dairy herds while reducing N pollution (Chen et al., 2021). Management and
70 breeding systems aimed at increasing NUE in dairy cows primarily focus on reducing urinary
71 N through proxies such as milk urea concentration (**MU**) or N (**MUN**) (Kauffman and St-Pierre,
72 2001; Spek et al., 2013; Bobbo et al., 2020; Ma et al., 2023). Researchers in New Zealand have
73 recently investigated the potential consequences of using MUN in genetic selection programs
74 to reduce N pollution in grazing dairy cows (Correa-Luna et al., 2021; Marshall et al., 2021,
75 2022). Marshall et al. (2021, 2022) demonstrated that selection for low MUN breeding values
76 can reduce urine N in cattle, while Correa-Luna et al. (2021) reported the opposite findings. In
77 the Walloon Region of Belgium, nearly all cattle graze on pasture from April to September
78 (Soyeurt et al., 2022), and farmers seek dairy cows that are best suited for grazing to minimize
79 feed production costs (Lefèvre et al., 2022). However, the oversupply of N for dairy cows is
80 not solely responsible for urinary N, but the imbalance between dietary N and energy in the
81 grazed pasture (Kebreab et al., 2002).

82 A recent large international collaboration revealed that predicting N excretion in dairy
83 cows requires the inclusion of N intake (**NINT**) and MUN to more accurately predict N
84 pollution, with the need for region-specific models (Bougouin et al., 2022). We proposed a
85 novel N efficiency index (**NEI**) that combines the estimated breeding values (**EBV**) of NINT,
86 milk true protein N (**MTPN**), and MUN yield (**MUNY**) through a selection index, considering

87 both NUE and N pollution simultaneously (Chen et al., 2022). MUNY is proportional to urinary
88 N excretion (Wisconsin Alumni Research Foundation, 2015). The NEI aims to maintain a
89 constant NINT, increase MTPN, and decrease MUNY through genetic selection (Chen et al.,
90 2022). However, the index was developed without incorporating genomic information.
91 Genomic information is valuable for selecting low-heritability traits, and is widely employed
92 in animal breeding programs. Currently, the single-step genomic best linear unbiased prediction
93 (**ssGBLUP**) is the most popular method for genetic evaluations of populations with both
94 genotyped and non-genotyped animals (Bermann et al., 2022).

95 To the best of our knowledge, genomic prediction is currently performed solely for
96 individual traits and not for indices. Results from genome-wide association analysis of NEI and
97 its composition indicated that NEI may reflect the interactions of its component traits (Chen et
98 al., 2023). The objectives of this study were to validate the genomic prediction of NEI and its
99 composition traits, and investigate the relationship between the single nucleotide
100 polymorphisms (**SNP**) effects estimated directly from NEI and indirectly from its composition
101 traits. This approach aims to verify whether NEI can be genomically predicted directly and
102 whether the index reflects its composition at the SNP level.

103 **Materials and Methods**

104 The study framework is shown in Figure 1.

105 ***Data***

106 **Phenotypic Data.** As an exploratory study, the data between 2012 and 2019 were selected from
107 the official milk recording database in the Walloon Region of Belgium. During this study's data
108 collection period, Holstein cows primarily grazed from April to September each year, while
109 their diet consisted of a total mixed ration (**TMR**, silage dominated) for the rest of the year. The
110 milk samples were analyzed by mid-infrared (**MIR**) spectrometry (commercial instruments
111 from FOSS) to assess milk composition (including MU) and generate MIR spectra. All milk

112 MIR spectra were standardized according to the method described by Grelet et al. (2015). The
113 3 features including NINT, MTPN, and MUNY were used in this study. These features were
114 divided into 2 classes: primiparous and multiparous (2 to 5 lactations) denoted as **NINT1**,
115 **MTPN1**, **MUNY1**, **NINT2+**, **MTPN2+**, and **MUNY2+** (six traits in total). The formulas used
116 to compute MTPN and MUNY were as follows: $MTPN = [(Milk\ yield \times Protein\ percent) / 6.38]$
117 $- MUNY]$ and $MUNY = [(Milk\ urea\ concentration) / 2.14] \times Milk\ yield]$ (WHO and FAO,
118 2011). The NINT of each cow was predicted using the equation developed by Grelet et al.
119 (2020). The determination coefficients and root mean square errors of validation for the NINT
120 equation were 0.71 and 0.07 kg/d, respectively.

121 The data editing procedure used for all the included traits was the same as explained by
122 Chen et al. (2022). Briefly, all records were restricted to the first 50 days in milk (**DIM**) because
123 of the predicament of the NINT model. The NINT model was built based on data from the first
124 50 DIM only. For NINT, two criteria were considered in the filtering procedure: 1) the
125 standardized Mahalanobis distance of the new predicted MIR spectra and the calibration dataset
126 is ≤ 3 ; 2) the predicted value of NINT was restricted within the range of ± 3 standard deviations
127 of the mean. Finally, 132,899 records, observed between 2012 and 2019 from 52,064 cows in
128 773 herds, were kept. The pedigree related to the dataset comprised 122,368 animals.

129 **Genotypic data.** The genotype data of 4,514 animals included in the pedigree were extracted
130 from the cattle genetic evaluation system in the Walloon Region of Belgium. The used chip
131 versions were BovineSNP50 K v1 to v3 (Illumina, San Diego, CA, USA). The SNPs common
132 between all three chips were kept. Non-mapped SNPs, SNP located on sex chromosomes, and
133 triallelic SNPs were excluded. A minimum GenCall Score of 0.15 and a minimum GenTrain
134 Score of 0.55 were used to keep SNP (Wilmot et al., 2022). Then, genotypes were imputed to
135 a high density (**HD**) SNP array using FImpute V2.2 software (Sargolzaei et al., 2014). The
136 SNPs with Mendelian conflicts, and those with minor allele frequency less than 5% were

137 excluded. The difference between observed heterozygosity and that expected under Hardy-
 138 Weinberg equilibrium was estimated, and if the difference was greater than 0.15, the SNP was
 139 excluded (Wiggans et al., 2009). Finally, 566,294 out of 730,539 SNPs, distributed on 29 *Bos*
 140 *taurus* autosomes, were kept.

141 *(Co)variance Component Estimation*

142 A six-trait (six traits; 3 features, 2 parity classes) repeatability model was used to estimate
 143 the variance components. For more detailed information on the model, refer to Chen et al.
 144 (2022). In summary, the model incorporated fixed effects such as herd-year-season of calving,
 145 standardized DIM and its quadratic term, and standardized calving age with constant, linear,
 146 and quadratic regression (nested within parities). Random effects included non-genetic cow
 147 effects, non-genetic cow effects specific to parity (for multiparous traits only), additive animal
 148 genetic effects, and residual effects. To calculate the relationship between animals, either a
 149 single (**H**) or pedigree-based (**A**) relationship matrix was employed. The **H** matrix combined
 150 the **A** and genomic (**G**)-based relationship matrices. The inverse of **H**, as defined by Aguilar et
 151 al. (2010) is as follows:

$$152 \quad \mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

153 where **A** is the numerator relationship matrix for all animals included in the pedigree; **A**₂₂ is the
 154 numerator relationship matrix for genotyped animals; **G** is the genomic relationship matrix
 155 obtained using the function described by VanRaden (2008). In addition, the inverse of all
 156 matrices considers the coefficient of inbreeding between individuals (Lourenco et al., 2020).

157 Computations were performed using the BLUPF90 family of programs (Misztal et al.,
 158 2014). The (co)variance components for NINT, MTPN, and MUNY were estimated by Gibbs
 159 sampling using the models described by Chen et al. (2022). Posterior means of (co)variances,
 160 heritabilities (**h**²), repeatabilities, genetic and phenotypic correlations were obtained using a

161 single chain of 130,000 iterations after a burn-in of 10,000 replicates. The formulas used to
 162 calculate the later four parameters have been previously described by Chen et al. (2021).
 163 Posterior convergence and approximated standard errors of (co)variances and four parameters
 164 were analyzed by POSTGIBBS90 (version 3.14). Repeatability, in this context, refers to within
 165 the six traits. For the traits in multiparous, repeatability was equal to the sum of the variances
 166 of the 3 random effects (non-genetic cow + non-genetic cow × parity + additive animal genetic)
 167 divided by the sum of the variances of all random effects.

168 *Estimated Breeding Values and Nitrogen Efficiency Index*

169 By using the estimated genetic variance components, the EBV and genomic EBV (**GEBV**)
 170 were estimated for the six studied traits through best linear unbiased prediction (**BLUP**) and
 171 ssGBLUP, respectively. Two scenarios were considered for estimating the (co)variance
 172 components: one with both pedigree, phenotypic, and genotypic data, and the other with only
 173 pedigree and phenotypic data. The EBV and GEBV for the six studied traits were estimated
 174 using the BLUPF90 program (version 1.71) with the same model used for (co)variance
 175 components estimation (Chen et al., 2022). The reliabilities (**REL**) of EBV and GEBV for the
 176 studied traits based upon the whole data were calculated by the formula:

$$177 \quad \text{REL}_j = 1 - \frac{\text{SE}_j^2}{\sigma_{uj}^2(1 + f)}$$

178 where REL_j is the reliability of j trait, SE_j is the standard error of prediction for j trait retrieved
 179 from the inverse of the left-hand side matrix of the mixed model equations, σ_{uj}^2 is the additive
 180 genetic variance, and f is the mean inbreeding coefficient for all animals included in the
 181 pedigree.

182 The calculation methods of the NEI and its reliability were the same as those described by
 183 Chen et al. (2022). In brief, the NEI was defined as a combination of the EBV or GEBV of
 184 NINT, MTPN, and MUNY through a selection index. The index weights (**a**) of the six studied

185 traits were calculated by selection responses, which assumed that the genetic selection
 186 responses for NINT, MTPN, and MUNY were 0, 1, and -1, respectively. This assumption
 187 allows genetic selection for NEI to maintain NINT while increasing MTPN and decreasing
 188 MUNY. The REL of NEI was estimated using the approach described in VanRaden et al.
 189 (2018):

$$190 \quad \mathbf{REL}_{\text{NEI}} = \frac{\mathbf{wG}_c\mathbf{w}'}{\mathbf{aG}_c\mathbf{a}'}$$

191 where $\mathbf{REL}_{\text{NEI}}$ was the REL of NEI, \mathbf{w} was obtained by multiplying the elements of \mathbf{a} by the
 192 square root of reliability for the EBV or GEBV of the six studied traits, \mathbf{G}_c was the genetic
 193 correlation matrix between the six studied traits.

194 *Nitrogen Efficiency Index and its Composition Traits Validation*

195 The theoretical accuracies of NEI and its composition traits in the whole dataset were
 196 calculated and were considered as the maximum accuracy of genetic selection in this dataset.
 197 The theoretical accuracy is equal to the square root of REL for NEI and each trait.

198 To assess the prediction accuracy of the genomic prediction in young animals, a linear
 199 regression-based method developed by Legarra and Reverter (2018) was used. The basic step
 200 of the linear regression method involves calculating the evaluation metrics by regressing the
 201 breeding value of the partial dataset according to the breeding value of the whole dataset. The
 202 data from 2017 to 2019 were set as missing values (n=38,906) and called the partial dataset.
 203 Since the whole dataset was small, only the records of the last three years were removed (2012
 204 to 2019). Both variance components and breeding values need to be estimated again in the
 205 partial dataset.

206 The validation population (called the focal individuals) consisted of cows and bulls (Figure
 207 1). Cows (n=4,413) born after 2015 were used as the focal individuals (without phenotypes in
 208 the partial dataset), of which 181 cows were genotyped. We verified the cows with and without

209 genotypic data, separately. In the case of the genotyped cows, BLUP validation was also
 210 performed to show whether differences in results could be caused by different validation
 211 populations. Bulls (n=55) were selected as the validation population under two criteria, of
 212 which 35 bulls were genotyped. The first criterion was used on bulls which only have daughters
 213 born after 2015 with records of six traits. The second criterion was used on those bulls with at
 214 least 5 daughters or at least 1 genotyped daughter. These 55 bulls altogether have 695 daughters
 215 with 872 records in this dataset. Since a small number of bulls met both the above criteria, bulls
 216 with and without genotypes were put together.

217 Four following metrics were used to measure prediction validation results in this study.

218 The population bias (μ_{wp}) was expected to be 0 if the evaluation was unbiased, which was
 219 defined as the following:

$$220 \quad \mu_{wp} = \overline{\hat{u}_p} - \overline{\hat{u}_w}$$

221 where $\overline{\hat{u}_p}$ and $\overline{\hat{u}_w}$ were average (G)EBV of focal individuals in the partial and whole datasets,
 222 respectively.

223 The dispersion (b_{wp}) was expected to be 1 if the evaluation showed no dispersion, which was
 224 defined as the following:

$$225 \quad b_{wp} = \frac{cov(\hat{u}_p, \hat{u}_w)}{var(\hat{u}_p)}$$

226 where \hat{u}_p and \hat{u}_w were EBV (or GEBV) of focal individuals in the partial and whole datasets,
 227 respectively.

228 The prediction accuracy (\widehat{acc}) was expected to be 1 if the evaluation was perfect. The
 229 predictive accuracy of the breeding values of focal individuals in the partial dataset was defined
 230 as the following:

$$231 \quad \widehat{acc} = \sqrt{\frac{cov(\hat{u}_p, \hat{u}_w)}{(1 - \bar{f})var(\hat{u}_p)}}$$

232 where \bar{f} was the average inbreeding coefficient of the focal individuals in the partial and whole
 233 datasets, respectively. Other parameters were the same as described above.

234 The relative increased accuracy with the phenotype (inc_{phen} ; Bermann et al., 2021) is
 235 expected to be low value, which means more stability between the partial and whole datasets.

236 The inc_{phen} was defined below:

$$237 \quad inc_{phen} = \frac{\widehat{acc}_w - \widehat{acc}_p}{\widehat{acc}_p} = \frac{1}{cor(\hat{u}_p, \hat{u}_w)} - 1$$

238 where \widehat{acc}_w and \widehat{acc}_p were accuracy in the whole and partial datasets, respectively. Other
 239 elements were the same as described above.

240 ***Relationship between Nitrogen Efficiency Index and its composition traits***

241 The Pearson correlation coefficients between the direct SNP effects estimated from NEI
 242 and indirect SNP effects estimated from its composition traits were estimated. The SNP effect
 243 ($\hat{\mathbf{a}}$) for the NEI and its composition traits were estimated using the POSTGSF90 software
 244 (version 1.73; Aguilar et al., 2014). The formula for the $\hat{\mathbf{a}}$ is the same as that described by Wang
 245 et al. (2012) but without iteration. The formula of $\hat{\mathbf{a}}$ was as follows:

$$246 \quad \hat{\mathbf{a}} = \mathbf{DZ}'_g[\mathbf{Z}_g\mathbf{DZ}'_g]^{-1}\hat{\mathbf{u}}$$

247 where $\mathbf{D} = \mathbf{I}$, means that the weight for all SNPs is 1; \mathbf{Z}_g was an incidence matrix of genotype
 248 for each SNP; $\hat{\mathbf{u}}$ was a vector of GEBV of each trait for genotyped animals (n=4,514) in the
 249 whole dataset.

250 The SNP effects for the six studied traits were multiplied by \mathbf{a} (relative weight) to calculate
 251 the SNP effects of NEI_hat. Then, the Pearson correlations between the SNP effects of NEI_hat
 252 and the SNP effects estimated directly from NEI were calculated.

253 **Results and Discussion**

254 ***Genetic Parameter***

255 The mean, h^2 , and repeatability for the six studied traits are shown in Table 1. The average
256 NUE, calculated as the mean MTPN divided by mean NINT, was lower for primiparous
257 (30.95%) than for multiparous (34.69%). This difference could be partly attributed to the more
258 developed rumen in multiparous cows. The average NUE values were within the normal range
259 reported by previous studies (Cantalapiedra-Hijar et al., 2018; Spanghero and Kowalski, 2021).
260 The h^2 and repeatability for the six traits ranged from 0.10 to 0.14, and 0.40 to 0.67,
261 respectively. These results are similar to our previous findings without using genotypic data
262 (Chen et al., 2022).

263 Table 2 shows the genetic and phenotypic correlations among the six studied traits. The
264 same feature in primiparous and multiparous showed high genetic correlations, ranging from
265 0.85 to 0.89. The NINT showed medium positive genetic correlations with MTPN (0.45 to
266 0.53), but low negative genetic correlations with MUNY (-0.08 to -0.20). These results suggest
267 that NINT may have a greater genetic influence on MTPN, but less on MUNY. Medium
268 positive genetic correlations were found between MTPN and MUNY (0.39 to 0.56). Our recent
269 study (Chen et al., 2023) indicated that the genomic regions affecting MTPN and MUNY were
270 not the same, supporting our hypothesis of simultaneous increases in MTPN and decreases in
271 MUNY through genetic selection. Phenotypic correlations were generally lower than genetic
272 correlations. Overall, the genetic and phenotypic correlations among the six traits are consistent
273 with our previous study (Chen et al., 2022). However, lower genetic correlations were estimated
274 between NINT and MUNY when genotypic data were used.

275 ***Validated Nitrogen Efficiency Index and its Composition Traits***

276 The mean theoretical accuracies of focal individuals from ssGBLUP were higher than
277 those from BLUP for both NEI and its composition traits (except MTPN₂₊ and MUNY₂₊ were

278 the same) (Table 3), which is consistent with Cesarani et al. (2021). The higher theoretical
279 accuracy of ssGBLUP is attributed to its smaller prediction error.

280 The validated results of the genomic prediction conducted for NEI and its composition
281 traits in genotyped cows, non-genotyped cows, and bulls through BLUP and ssGBLUP are
282 presented in Table 4.

283 First, the results of BLUP and ssGBLUP were compared. The estimated μ_{wp} (nearly 0)
284 and b_{wp} (nearly 1) for the NEI and its composition traits of BLUP and ssGBLUP in genotyped
285 cows indicate that bias and dispersion for genomic prediction between BLUP and ssGBLUP
286 are similar. Similar results for the NEI and its composition traits were found for non-genotyped
287 cows and bulls, except for the b_{wp} of bulls. Gao et al. (2021) observed similar results for feed
288 intake in pigs when comparing BLUP and ssGBLUP, even though they used more genotyped
289 animals than non-genotyped animals. As expected, the \widehat{acc} of the NEI and its composition traits
290 estimated by ssGBLUP in genotyped cows were higher than that estimated by BLUP, ranging
291 from 0.48 to 0.66. The \mathbf{H} relationship matrix, used in the ssGBLUP, may explain this finding
292 (Cesarani et al., 2021). Even for non-genotyped animals, the \widehat{acc} of the NEI and its composition
293 traits estimated by ssGBLUP were slightly higher than that estimated by BLUP, ranging from
294 0.36 to 0.55. Some of the genotyped animals in the dataset are related to the non-genotyped
295 animals which might partially explain this result. The \widehat{acc} of NEI and its composition traits
296 estimated by ssGBLUP in bulls were higher than that estimated by BLUP, ranging from 0.39
297 to 0.56. Similarly, multiple studies have shown that the accuracy of ssGBLUP is higher than
298 BLUP (Bermann et al., 2021; Cesarani et al., 2021). The inc_{phen} of the NEI and its composition
299 traits in genotyped cows and bulls estimated by ssGBLUP were lower than that estimated by
300 BLUP, which is in line with Bermann et al. (2021) study on chicken mortality. However, the
301 inc_{phen} of NEI and its composition traits in non-genotyped animals estimated by ssGBLUP

302 and BLUP were similar, except for NINT. This suggests that when making genomic predictions
303 for non-phenotyped cows, providing their own genotypic data is a good predictor.

304 On the other hand, the obtained results were compared between genotyped cows, non-
305 genotyped cows, and bulls. The μ_{wp} of NINT and MTPN estimated by ssGBLUP (or BLUP)
306 for genotyped cows, non-genotyped cows, and bulls were similar, while the μ_{wp} of MUNY and
307 NEI estimated by ssGBLUP (or BLUP) for genotyped cows and bulls were higher than that
308 estimated for non-genotyped cows. The b_{wp} of NEI and its composition traits estimated by
309 ssGBLUP (or BLUP) were similar between genotyped and non-genotyped cows. However, the
310 b_{wp} of NEI and its composition traits estimated by ssGBLUP (or BLUP) in bulls were worse
311 than in cows. The \widehat{acc} of the NEI and its composition traits estimated by ssGBLUP in
312 genotyped cows were higher than that in non-genotyped cows and bulls, which can be a
313 consequence of using extra information available (SNP data) for genotyped cows. This is an
314 advantage of using genomic selection for low heritability traits (Parker Gaddis et al., 2014).
315 However, the \widehat{acc} of the NEI and its composition traits estimated by BLUP in genotyped cows
316 and bulls (after removing genotypic data) were lower than that in non-genotyped cows, which
317 can be attributed to the smaller number of cows in the genotyped group ($n = 181$) and bulls
318 ($n=55$) compared with non-genotyped ($n = 4,232$). The inc_{phen} of NEI and its composition
319 traits estimated by ssGBLUP (or BLUP) ranged from 17.43% (32.14%) to 31.42% (76.17%) in
320 genotyped cows, from 23.14% (28.06%) to 57.38% (60.44%) in non-genotyped cows, and from
321 27.93% (52.27%) to 65.73% (86.07%) in bulls.

322 In short, these findings showed that genomic information is beneficial for genomic
323 predictions for NEI and its composition traits. The highest mean prediction accuracies for NEI
324 and its composition traits were estimated for genotyped cows using ssGBLUP. Furthermore, it
325 is worth mentioning that even for non-genotyped and non-phenotyped animals in the partial

326 dataset, the prediction accuracy of genomic prediction for NEI was nearly the theoretical
327 accuracy by ssGBLUP (0.52 vs 0.58).

328 *Relationship between Nitrogen Efficiency Index and its Composition Traits*

329 Figure 2 shows the 566,294 SNPs effects values of the NEI, ranging from -0.003 to 0.003.
330 The Pearson correlation between SNP effects estimated directly from NEI and indirectly from
331 its composition traits was 1 with an intercept of 0 (Figure 2), which indicates that indirect
332 genomic prediction can be performed on the NEI index.

333 Over the past few decades, the breeding goal of dairy cattle has gradually shifted from
334 increasing milk production to balanced breeding (Miglior et al., 2017), which has resulted in
335 more traits being added to the breeding system. With increasing the number of traits in a
336 breeding program, breeders usually classify the traits to generate an index, and then combine
337 them to generate a total index for breeding. Our results showed that the SNP effects estimated
338 directly from the generated index are completely consistent with those estimated from its
339 composition traits (Figure 2). This implies that if SNP effects of the total index are obtained, it
340 becomes more convenient to perform genomic prediction of the total index on young bulls.
341 Indirect genomic prediction offers the advantage of reducing the computational cost of large-
342 scale genomic selection (Tsuruta et al., 2021), thus facilitating the implementation of genomic
343 prediction at the national level.

344 Another potential advantage of indirect genomic prediction of NEI is the potential
345 reduction in prediction error. There are potential errors in the calculation of breeding values for
346 each of the six examined traits, which may lead to a reduction in the accuracy of NEI. The
347 potential improvement in the accuracy of the NEI can be achieved by reducing the number of
348 steps in its calculation. However, it is important to note that the theoretical accuracy of indirect
349 genomic predictions of NEI needs further review, as conducted by Garcia et al. (2022) for post-

350 weaning weight gain. Furthermore, the effectiveness of indirect genomic prediction for indices
351 should be validated on more indices to ensure its applicability.

352 *Limitations and Perspectives*

353 The results of this study also require several points of attention. First, the NEI is a proxy
354 for the NUE of dairy cattle, therefore, it is not identical to the genetic merit of NUE. Moreover,
355 there were some prediction errors in the NEI calculation process, such as the prediction models
356 of the MIR-based phenotypes and the estimation of relevant GEBV. To address these issues,
357 two approaches (data edited) were adopted to mitigate potential problems. Additional reference
358 phenotypes are expected to be added for improving the prediction models of the MIR-based
359 phenotypes. Furthermore, Tiplady et al. (2022) recently reported that genetic correlations
360 between directly measured and MIR-predicted fatty acid and protein fractions were typically
361 high. This suggests that MIR-based phenotypes represent a valid estimation of genetic
362 contributions to the directly measured traits. The estimation of relevant GEBV will become
363 accurate with an increasing number of genotyped animals.

364 This study should be considered an exploration study, also based on a limited dataset. This
365 implies that before constructing a national-level NEI combining genomic information and its
366 integration into a global breeding goal additional work will be required. Although the mean
367 prediction accuracies (0.52-0.59) of NEI in focal individuals from ssGBLUP are not very high,
368 they are not too far from mean theoretical accuracies (0.58-0.66). The average reliability (square
369 of theoretical accuracy) of the bulls ($n = 736$) was 0.68 when establishing NEI without genomic
370 information (Chen et al., 2022), indicating that NEI can still perform genomic predictions when
371 sufficient records are available, especially when combined with genomic information.

372 The small number of records extracted from the official milk recording database is the
373 main reason for the low accuracy obtained. Especially, NINT data used was still limited to the
374 first 50 DIM because the currently available MIR prediction equation is established using only

375 data from the first 50 DIM. However, extending it to the entire lactation period is planned when
376 relevant reference data are available. This will be our focus as the next step, which should also
377 increase the reliability of NEI because we could use MIR-predicted phenotypes across the
378 whole lactation. Another possible way to increase the reliability of NEI is to increase the number
379 of genotyped animals. The number of cows genotyped has been increasing rapidly in recent
380 years (Wiggans and Carrillo, 2022), not only in the USA but also to a lesser degree in the
381 Walloon Region of Belgium.

382 To implement the routine use of NEI, its integration into a broader breeding goal would
383 be crucial. In the context of the Walloon genetic evaluation system, which defines sub-indexes
384 (Vanderick et al., 2022), utilizing NEI as an N efficiency sub-index or part of a broader
385 efficiency sub-index would be relevant. This approach could serve as a starting point for
386 breeding efforts aimed at improving NUE in cattle.

387 **Conclusions**

388 This study demonstrated that genomic prediction can be successfully applied to the NEI,
389 providing valuable insights into NUE in dairy cattle. Even when animals in the partial dataset
390 lacked both genotypic and phenotypic data, the genomic prediction for NEI achieved a mean
391 prediction accuracy close to the theoretical accuracy by ssGBLUP (0.52 vs 0.58). Moreover,
392 we explored the potential benefits of both direct and indirect genomic prediction for the NEI
393 index, which could prove advantageous for larger datasets at the national level. This study also
394 confirmed that NEI effectively reflects its composition at the SNP level. However, it is
395 important to acknowledge that further validation in a larger dataset is needed to bolster the
396 credibility of our findings. By adding genomic information to the establishment of NEI, this
397 study lays the groundwork for future routine genomic evaluation programs. The integration of
398 NEI into such programs holds promise for enhancing N efficiency breeding efforts in dairy

399 cattle, thus promoting more sustainable and environmentally responsible practices in the dairy
400 industry.

401 **Supplementary Materials**

402 The R code for calculating the NEI and its reliability with example data can be found on
403 GitHub (https://github.com/Yansen0515/GP_NEI)

404 **Acknowledgments**

405 The China Scholarship Council (Beijing) is acknowledged for funding the PhD project of
406 Yansen Chen and Hongqing Hu. The authors would like to thank the European Union's Seventh
407 Framework Program for funding the GplusE project under grant agreement n° 613689. The
408 views expressed in this publication are the sole responsibility of the authors and do not
409 necessarily reflect the views of the European Commission. The authors acknowledge the
410 support of the Walloon Government (Service Public de Wallonie – Direction Générale
411 Opérationnelle Agriculture, Ressources Naturelles et Environnement, SPW-DGARNE; Namur,
412 Belgium) and the use of the computation resources of the University of Liège – Gembloux
413 Agro-Bio Tech (Gembloux, Belgium) provided by the technical platform Calcul et
414 Modélisation Informatique (CAMI) of the TERRA Teaching and Research Centre, partly
415 supported by the National Fund for Scientific Research (F.R.S.-FNRS, Brussels, Belgium)
416 under Grants No. T.0095.19 (PDR “DEEPSELECT”) and J.0174.18 (CDR “PREDICT-2”). The
417 authors have not stated any conflicts of interest.

418

419 **References**

420 Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic:
421 A unified approach to utilize phenotypic, full pedigree, and genomic information for

- 422 genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.
423 <https://doi.org/10.3168/jds.2009-2730>.
- 424 Aguilar, I., I. Misztal, S. Tsuruta, A. Legarra, and H. Wang. 2014. PREGSF90–POSTGSF90:
425 Computational tools for the implementation of single-step genomic selection and
426 genome-wide association with ungenotyped individuals in BLUPF90 programs.
427 Proceedings of the 10th world congress of genetics applied to livestock production,
428 Vancouver, BC, Canada. Accessed August 21, 2022.
429 [http://www.ainfo.inia.uy/digital/bitstream/item/15445/1/Aguilar-et-al.-2014.-](http://www.ainfo.inia.uy/digital/bitstream/item/15445/1/Aguilar-et-al.-2014.-WCGALP.pdf)
430 [WCGALP.pdf](http://www.ainfo.inia.uy/digital/bitstream/item/15445/1/Aguilar-et-al.-2014.-WCGALP.pdf)
- 431 Bermann, M., A. Cesarani, I. Misztal, and D. Lourenco. 2022. Past, present, and future
432 developments in single-step genomic models. *Ital. J. Anim. Sci.* 21:673–685.
433 <https://doi.org/10.1080/1828051X.2022.2053366>.
- 434 Bermann, M., A. Legarra, M.K. Hollifield, Y. Masuda, D. Lourenco, and I. Misztal. 2021.
435 Validation of single-step GBLUP genomic predictions from threshold models using the
436 linear regression method: An application in chicken mortality. *J. Anim. Breed. Genet.*
437 138:4–13. <https://doi.org/10.1111/jbg.12507>.
- 438 Bobbo, T., M. Penasa, A. Rossoni, and M. Cassandro. 2020. Short communication: Genetic
439 aspects of milk urea nitrogen and new indicators of nitrogen efficiency in dairy cows. *J.*
440 *Dairy Sci.* 103:9207–9212. <https://doi.org/10.3168/jds.2020-18445>.
- 441 Bougouin, A., A. Hristov, J. Dijkstra, M.J. Aguerre, S. Ahvenjärvi, C. Arndt, A. Bannink, A.R.
442 Bayat, C. Benchaar, T. Boland, W.E. Brown, L.A. Crompton, F. Dehareng, I. Dufrasne,
443 M. Eugène, E. Froidmont, S. van Gastelen, P.C. Garnsworthy, A. Halmemies-Beauchet-
444 Filleau, S. Herremans, P. Huhtanen, M. Johansen, A. Kidane, M. Kreuzer, B. Kuhla, F.
445 Lessire, P. Lund, E.M.K. Minnée, C. Muñoz, M. Niu, P. Nozière, D. Pacheco, E.
446 Prestløkken, C.K. Reynolds, A. Schwarm, J.W. Spek, M. Terranova, A. Vanhatalo, M.A.

- 447 Wattiaux, M.R. Weisbjerg, D.R. Yáñez-Ruiz, Z. Yu, and E. Kebreab. 2022. Prediction of
448 nitrogen excretion from data on dairy cows fed a wide range of diets compiled in an
449 intercontinental database: A meta-analysis. *J. Dairy Sci.* 105:7462–7481.
450 <https://doi.org/10.3168/jds.2021-20885>.
- 451 Cantalapiedra-Hijar, G., R.J. Dewhurst, L. Cheng, A.R.J. Cabrita, A.J.M. Fonseca, P. Nozière,
452 D. Makowski, H. Fouillet, and I. Ortigues-Marty. 2018. Nitrogen isotopic fractionation
453 as a biomarker for nitrogen use efficiency in ruminants: A meta-analysis. *Animal*
454 12:1827–1837. <https://doi.org/10.1017/S1751731117003391>.
- 455 Cesarani, A., A. Garcia, J. Hidalgo, L. Degano, D. Vicario, N.P.P. Macciotta, and D. Lourenco.
456 2021. Genomic information allows for more accurate breeding values for milkability in
457 dual-purpose Italian Simmental cattle. *J. Dairy Sci.* 104:5719–5727.
458 <https://doi.org/10.3168/jds.2020-19838>.
- 459 Chen, Y., H. Atashi, C. Grelet, R.R. Mota, S. Vanderick, H., Hu, GplusE Consortium, and N.
460 Gengler. 2023. Genome-wide association study and functional annotation analyses for
461 nitrogen efficiency index and its composition traits in dairy cattle. *J. Dairy Sci.* 106:3397–
462 3410. <https://doi.org/10.3168/jds.2022-22351>.
- 463 Chen, Y., H. Atashi, C. Grelet, S. Vanderick, H. Hu, and N. Gengler. 2022. Defining a nitrogen
464 efficiency index in Holstein cows and assessing its potential impact on the breeding
465 program of bulls. *J. Dairy Sci.* 105: 7575-7587. <https://doi.org/10.3168/jds.2021-21681>.
- 466 Chen, Y., S. Vanderick, R.R. Mota, C. Grelet, GplusE Consortium, and N. Gengler. 2021.
467 Estimation of genetic parameters for predicted nitrogen use efficiency and losses in early
468 lactation of Holstein cows. *J. Dairy Sci.* 104:4412-4423.
469 <https://doi.org/10.3168/jds.2020-18849>.

- 470 Correa-Luna, M., D. Donaghy, P. Kemp, M. Schutz, and N. López-Villalobos. 2021. Nitrogen
471 use efficiency and excretion in grazing cows with high and low milk urea nitrogen
472 breeding values. *Sustainability* 13:1-12. <https://doi.org/10.3390/su13179827>.
- 473 Gao, H., G. Su, J. Jensen, P. Madsen, O.F. Christensen, B. Ask, B.G. Poulsen, T. Ostensen, and
474 B. Nielsen. 2021. Genetic parameters and genomic prediction for feed intake recorded at
475 the group and individual level in different production systems for growing pigs. *Genet.*
476 *Sel. Evol.* 53:1–12. <https://doi.org/10.1186/s12711-021-00624-3>.
- 477 Garcia, A., I. Aguilar, A. Legarra, S. Tsuruta, I. Misztal, and D. Lourenco. 2022. Theoretical
478 accuracy for indirect predictions based on SNP effects from single-step GBLUP. *Genet.*
479 *Sel. Evol.* 54:1–11. <https://doi.org/10.1186/s12711-022-00752-4>.
- 480 Grelet, C., J.A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015.
481 Standardization of milk mid-infrared spectra from a European dairy network. *J. Dairy*
482 *Sci.* 98:2150–2160. <https://doi.org/10.3168/jds.2014-8764>.
- 483 Grelet, C., E. Froidmont, L. Foldager, M. Salavati, M. Hostens, C.P. Ferris, K.L. Ingvarsen,
484 M.A. Crowe, M.T. Sorensen, J.A. Fernandez Pierna, A. Vanlierde, N. Gengler, GplusE
485 Consortium, and F. Dehareng. 2020. Potential of milk mid-infrared spectra to predict
486 nitrogen use efficiency of individual dairy cows in early lactation. *J. Dairy Sci.* 103:4435–
487 4445. <https://doi.org/10.3168/jds.2019-17910>.
- 488 Kauffman, A.J., and N.R. St-Pierre. 2001. The relationship of milk urea nitrogen to urine
489 nitrogen excretion in Holstein and Jersey cows. *J. Dairy Sci.* 84:2284–2294.
490 [https://doi.org/10.3168/jds.S0022-0302\(01\)74675-9](https://doi.org/10.3168/jds.S0022-0302(01)74675-9).
- 491 Kebreab, E., J. France, J.A.N. Mills, R. Allison, and J. Dijkstra. 2002. A dynamic model of N
492 metabolism in the lactating dairy cow and an assessment of impact of N excretion on the
493 environment. *J. Anim. Sci.* 80:248–259. <https://doi.org/10.2527/2002.801248x>.

- 494 Lefèvre, A., V. Decruyenaere, and E. Froidmont. 2022. Transition towards more robust cows
495 on Walloon dairy farms for a better use of forage resources. Dair'Innov 2022 congress,
496 Namur, Belgium.
- 497 Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias
498 of predictions of breeding values and future phenotypes using the LR method. *Genet. Sel.*
499 *Evol.* 50:1–18. <https://doi.org/10.1186/s12711-018-0426-6>.
- 500 Lourenco, D., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguilar, and I. Misztal. 2020. Single-step
501 genomic evaluations from theory to practice: using SNP chips and sequence data in
502 blupf90. *Genes (Basel)*. 11:1–32. <https://doi.org/10.3390/genes11070790>.
- 503 Ma, L., H. Luo, L.F. Brito, Y. Chang, Z. Chen, W. Lou, F. Zhang, L. Wang, G. Guo, and Y.
504 Wang. 2023. Estimation of genetic parameters and single-step genome-wide association
505 studies for milk urea nitrogen in Holstein cattle. *J. Dairy Sci.* 106:352–363.
506 <https://doi.org/10.3168/jds.2022-21857>.
- 507 Marshall, C.J., M.R. Beck, K. Garrett, G.K. Barrell, O. Al-Marashdeh, and P. Gregorini. 2021.
508 Nitrogen balance of dairy cows divergent for milk urea nitrogen consuming either
509 plantain or perennial ryegrass. *Animals* 11:1–18. <https://doi.org/10.3390/ani11082464>.
- 510 Marshall, C.J., M.R. Beck, K. Garrett, G.K. Barrell, O. Al-Marashdeh, and P. Gregorini. 2022.
511 Urine and fecal excretion patterns of dairy cows divergent for milk urea nitrogen breeding
512 values consuming either a plantain or ryegrass diet. *J. Dairy Sci.* 105:4218–4236.
513 <https://doi.org/10.3168/jds.2021-21490>.
- 514 Miglior, F., A. Fleming, F. Malchiodi, L.F. Brito, P. Martin, and C.F. Baes. 2017. A 100-Year
515 Review: Identification and genetic selection of economically important traits in dairy
516 cattle. *J. Dairy Sci.* 100:10251-10271. <https://doi.org/10.3168/jds.2017-12968>.

- 517 Misztal, I., S. Tsuruta, D. A. L. Lourenco, Y. Masuda, I. Aguilar, A. Legarra, and Z. Vitezica.
518 2014. Manual for BLUPF90 family programs. University of Georgia. Accessed Aug. 01,
519 2022. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.
- 520 Moran, D., and E. Wall. 2011. Livestock production and greenhouse gas emissions: Defining
521 the problem and specifying solutions. *Anim. Front.* 1:19–25.
522 <https://doi.org/10.2527/af.2011-0012>.
- 523 Parker Gaddis, K.L., J.B. Cole, J.S. Clay, and C. Maltecca. 2014. Genomic selection for
524 producer-recorded health event data in US dairy cattle. *J. Dairy Sci.* 97:3190–3199.
525 <https://doi.org/10.3168/jds.2013-7543>.
- 526 Rochette, P., M.H. Chantigny, N. Ziadi, D.A. Angers, G. Bélanger, É. Charbonneau, D.
527 Pellerin, C. Liang, and N. Bertrand. 2014. Soil nitrous oxide emissions after deposition
528 of dairy cow excreta in eastern Canada. *J. Environ. Qual.* 43:829–841.
529 <https://doi.org/10.2134/jeq2013.11.0474>
- 530 Sargolzaei, M., J.P. Chesnais, and F.S. Schenkel. 2014. A new approach for efficient genotype
531 imputation using information from relatives. *BMC Genom.* 15:1-12.
532 <https://doi.org/10.1186/1471-2164-15-478>.
- 533 Soyeurt, H., C. Gerards, C. Nickmilder, J. Bindelle, S. Franceschini, F. Dehareng, D. Veselko, C.
534 Bertozzi, N. Gengler, A. Marvuglia, A. Bayram, and A. Tedde. 2022. Prediction of indirect
535 indicators of a grass-based diet by milk Fourier Transform mid-infrared spectroscopy to assess
536 the feeding typologies of dairy farms. *Animals* 12:1-13. <https://doi.org/10.3390/ani12192663>.
- 537 Spanghero, M., and Z.M. Kowalski. 2021. Updating analysis of nitrogen balance experiments
538 in dairy cows. *J. Dairy Sci.* 104:7725–7737. <https://doi.org/10.3168/jds.2020-19656>.
- 539 Spek, J.W., J. Dijkstra, G. Van Duinkerken, and A. Bannink. 2013. A review of factors
540 influencing milk urea concentration and its relationship with urinary urea excretion in

- 541 lactating dairy cattle. *J. Agric. Sci.* 151:407–423.
 542 <https://doi.org/10.1017/S0021859612000561>.
- 543 Tiplady, K.M., T.J. Lopdell, R.G. Sherlock, T.J.J. Johnson, R.J. Spelman, B.L. Harris, S.R.
 544 Davis, M.D. Littlejohn, and D.J. Garrick. 2022. Comparison of the genetic characteristics
 545 of directly measured and Fourier-transform mid-infrared-predicted bovine milk fatty
 546 acids and proteins. *J. Dairy Sci.* In press. <https://doi.org/10.3168/jds.2022-22089>.
- 547 Tsuruta, S., D.A.L. Lourenco, Y. Masuda, T.J. Lawlor, and I. Misztal. 2021. Reducing
 548 computational cost of large-scale genomic evaluation by using indirect genomic
 549 prediction. *JDS Commun.* 2:356–360. <https://doi.org/10.3168/jdsc.2021-0097>.
- 550 Vanderick, S., R. R. Mota, K. Wijnrocs, and N. Gengler. 2022. Description of the genetic
 551 evaluation systems used in the Walloon Region of Belgium. Accessed November 21,
 552 2022. <http://www.elinfo.be/indexEN.html>.
- 553 VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.*
 554 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- 555 VanRaden, P.M., J.B. Cole, and K.L. Parker Gaddis. 2018. Net merit as a measure of lifetime
 556 profit: 2018 revision. Accessed Aug. 01, 2022.
 557 <https://www.aipl.arsusda.gov/reference/nmcalc-2018.htm>.
- 558 Wang, H., I. Misztal, I. Aguilar, A. Legarra, and W.M. Muir. 2012. Genome-wide association
 559 mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb.)*.
 560 94:73–83. <https://doi.org/10.1017/S0016672312000274>.
- 561 WHO and FAO (World Health Organization and Food and Agriculture Organization of the
 562 United Nations). 2011. *Codex Alimentarius: Milk and Milk Products*. 2nd ed. WHO
 563 FAO, Rome, Italy.
- 564 Wiggans, G.R., and J.A. Carrillo. 2022. Genomic selection in United States dairy cattle. *Front.*
 565 *Genet.* 13:1–7. <https://doi.org/10.3389/fgene.2022.994466>.

- 566 Wiggans, G.R., T.S. Sonstegard, P.M. VanRaden, L.K. Matukumalli, R.D. Schnabel, J.F.
567 Taylor, F.S. Schenkel, and C.P. van Tassell. 2009. Selection of single-nucleotide
568 polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the
569 United States and Canada. *J. Dairy Sci.* 92:3431–3436. [https://doi.org/10.3168/jds.2008-](https://doi.org/10.3168/jds.2008-1758)
570 [1758](https://doi.org/10.3168/jds.2008-1758).
- 571 Wisconsin Alumni Research Foundation. 2015. Milk Urea-N yield (MUNY) as a nutritional
572 and environmental management tool for the dairy industry. US Pat. No. 9,955,673 B1.
- 573 Wilmot, H., J. Bormann, H. Soyeurt, X. Hubin, G. Glorieux, P. Mayeres, C. Bertozzi, and N.
574 Gengler. 2022. Development of a genomic tool for breed assignment by comparison of
575 different classification models: Application to three local cattle breeds. *J. Anim. Breed.*
576 *Genet.* 139:40–61. <https://doi.org/10.1111/jbg.12643>.

577 **Table 1** Description statistics of the six studied traits¹ and their heritabilities and repeatabilities
 578 in the whole dataset

Parity class ²	NINT (kg/d)	MTPN (kg/d)	MUNY (g/d)
1 (n = 40,916)			
Mean	0.42	0.13	2.93
SD ³	0.06	0.03	1.08
Heritability	0.14 ± 0.01	0.13 ± 0.01	0.14 ± 0.01
Repeatability	0.45 ± 0.01	0.61 ± 0.00	0.40 ± 0.01
2+ (n = 91,938)			
Mean	0.49	0.17	3.74
SD	0.07	0.04	1.56
Heritability	0.13 ± 0.01	0.11 ± 0.01	0.10 ± 0.01
Repeatability	0.53 ± 0.00	0.67 ± 0.00	0.43 ± 0.00

579 ¹NINT – N intake; MTPN – Milk true protein N; MUNY – Milk urea N yield

580 ²Parity class: based on the parity, the data were divided into 2 classes including primiparous
 581 (class 1) and multiparous (class 2+)

582 ³SD: standard deviation

583 **Table 2.** Genetic correlations (above the diagonal) and phenotypic correlations (below the
 584 diagonal) are among the six studied traits.

Traits ¹	NINT1	MTPN1	MUNY1	NINT2+	MTPN2+	MUNY2+
NINT1		0.48 ± 0.04	-0.20 ± 0.05	0.87 ± 0.02	0.50 ± 0.04	-0.16 ± 0.05
MTPN1	0.36 ± 0.01		0.56 ± 0.04	0.53 ± 0.04	0.85 ± 0.03	0.46 ± 0.05
MUNY1	0.06 ± 0.01	0.37 ± 0.00 ²		-0.03 ± 0.04	0.39 ± 0.04	0.89 ± 0.03
NINT2+	0.17 ± 0.01	0.14 ± 0.01	0.02 ± 0.01		0.61 ± 0.03	-0.08 ± 0.05
MTPN2+	0.11 ± 0.01	0.20 ± 0.01	0.11 ± 0.01	0.43 ± 0.00		0.45 ± 0.04
MUNY2+	-0.02 ± 0.01	0.09 ± 0.01	0.16 ± 0.01	0.14 ± 0.00	0.41 ± 0.00	

585 ¹NINT1 - N intake in primiparous cows; MTPN1 – Milk true protein N in primiparous cows;
 586 MUNY1 - Milk urea N yield in primiparous cows; NINT2+ - N intake in multiparous cows;
 587 MTPN2+ - Milk true protein N in multiparous cows; MUNY2+ - Milk urea N yield in
 588 multiparous cows

589 ²standard error is less than 0.005.

590 **Table 3.** The mean theoretical accuracies in the whole dataset for genotyped cows, non-
 591 genotyped cows, and bulls of nitrogen efficiency index (NEI) and its composition traits¹

	NINT1	MTPN1	MUNY1	NINT2+	MTPN2+	MUNY2+	NEI
Genotyped cows (n=181) ²							
BLUP	0.63	0.58	0.61	0.61	0.57	0.59	0.57
ssGBLUP ³	0.72	0.67	0.69	0.71	0.67	0.69	0.66
Non-genotyped cows (n=4,232)							
BLUP	0.62	0.58	0.60	0.60	0.57	0.59	0.57
ssGBLUP	0.63	0.59	0.61	0.61	0.57	0.59	0.58
Bulls (n=55) ⁴							
BLUP	0.64	0.60	0.62	0.61	0.58	0.60	0.59
ssGBLUP	0.70	0.66	0.68	0.68	0.64	0.67	0.65

592 ¹NINT1 - N intake in primiparous cows; MTPN1 – Milk true protein N in primiparous cows;
 593 MUNY1 - Milk urea N yield in primiparous cows; NINT2+ - N intake in multiparous cows;
 594 MTPN2+ - Milk true protein N in multiparous cows; MUNY2+ - Milk urea N yield in
 595 multiparous cows

596 ²Ranges of standard deviation for genotype cows, non-genotyped cows, and bulls are 0.06 to
 597 0.10, 0.05 to 0.06, and 0.03 to 0.05, respectively

598 ³ssGBLUP - Single-step genomic BLUP

599 ⁴35 of 55 bulls had genotype

600 **Table 4.** Validated nitrogen efficiency index (NEI) and its composition traits¹ by linear
 601 regression

		NINT1	MTPN1	MUNY1	NINT2+	MTPN2+	MUNY2+	NEI
Genotyped cows (n=181)								
	bias	-0.01	0.00	-0.05	-0.01	0.00	-0.07	0.16
BLUP	dispersion	1.05	1.18	1.20	1.07	1.06	1.04	1.06
	accuracy	0.40	0.37	0.33	0.41	0.34	0.35	0.45
	inc _{phen} (%) ²	76.17	38.91	60.02	64.92	45.54	58.42	32.14
	bias	-0.01	0.00	-0.07	-0.01	0.00	-0.11	0.14
ssGBLUP ³	dispersion	1.06	0.97	1.13	0.99	0.93	0.99	1.02
	accuracy	0.63	0.48	0.49	0.66	0.51	0.50	0.59
	inc _{phen} (%)	25.30	30.92	31.42	23.45	22.59	30.11	17.43
Non-genotyped cows (n=4,232)								
	bias	0.00	0.00	-0.03	-0.01	0.00	-0.03	0.01
BLUP	dispersion	1.08	1.14	1.10	1.15	1.09	0.96	1.03
	accuracy	0.47	0.43	0.33	0.50	0.39	0.35	0.48
	inc _{phen} (%)	52.07	29.69	60.44	41.02	33.73	54.36	28.06
	bias	0.00	0.00	-0.04	-0.01	0.00	-0.05	0.00
ssGBLUP	dispersion	1.13	1.06	1.06	1.16	1.04	0.95	1.00
	accuracy	0.53	0.45	0.35	0.55	0.43	0.36	0.52
	inc _{phen} (%)	35.42	28.62	57.38	27.88	27.27	52.26	23.14
Bulls (n=55) ⁴								
	bias	0.00	0.00	-0.09	-0.01	0.00	-0.11	0.10
BLUP	dispersion	1.03	1.61	0.89	1.20	1.32	0.68	0.88
	accuracy	0.34	0.32	0.33	0.39	0.31	0.33	0.43
	inc _{phen} (%)	86.07	56.43	78.54	55.29	52.27	78.70	55.47
	bias	0.00	0.00	-0.11	-0.01	0.00	-0.14	0.06
ssGBLUP	dispersion	1.08	1.14	1.02	1.14	0.99	0.67	0.99
	accuracy	0.52	0.41	0.42	0.56	0.42	0.39	0.52
	inc _{phen} (%)	33.14	50.23	47.04	27.93	49.98	65.37	30.71

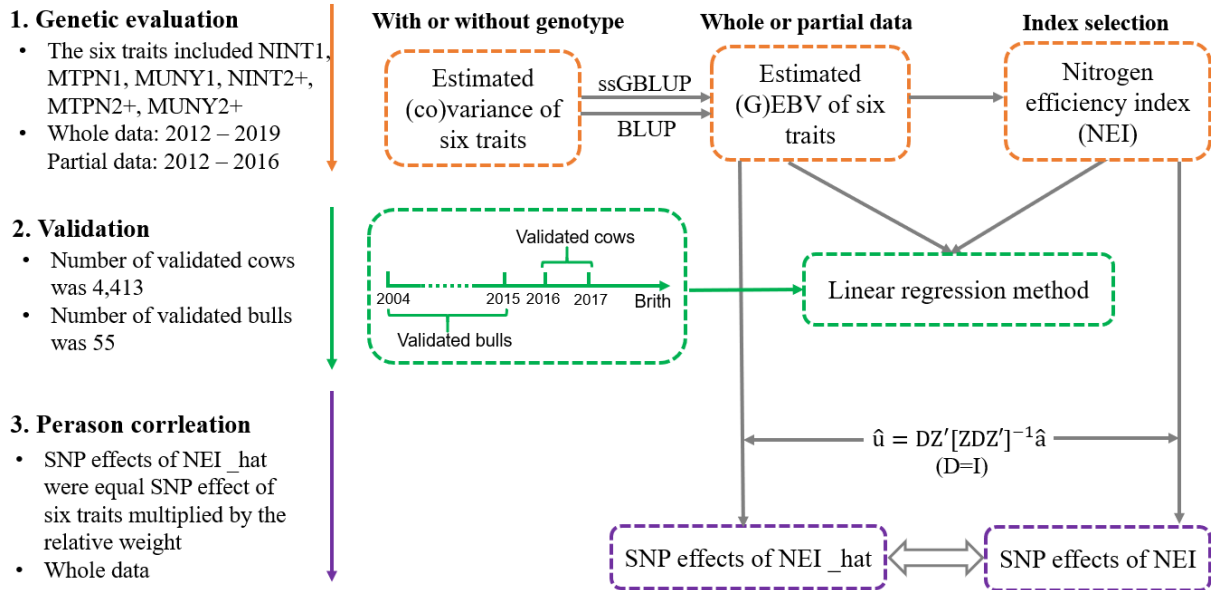
602 ¹NINT1 - N intake in primiparous cows; MTPN1 – Milk true protein N in primiparous cows;
 603 MUNY1 - Milk urea N yield in primiparous cows; NINT2+ - N intake in multiparous cows;
 604 MTPN2+ - Milk true protein N in multiparous cows; MUNY2+ - Milk urea N yield in
 605 multiparous cows

606 ²inc_{phen}(%) - Increase reliability when adding phenotypic

607 ³ssGBLUP - Single-step genomic BLUP

608 ⁴35 of 55 bulls had genotype

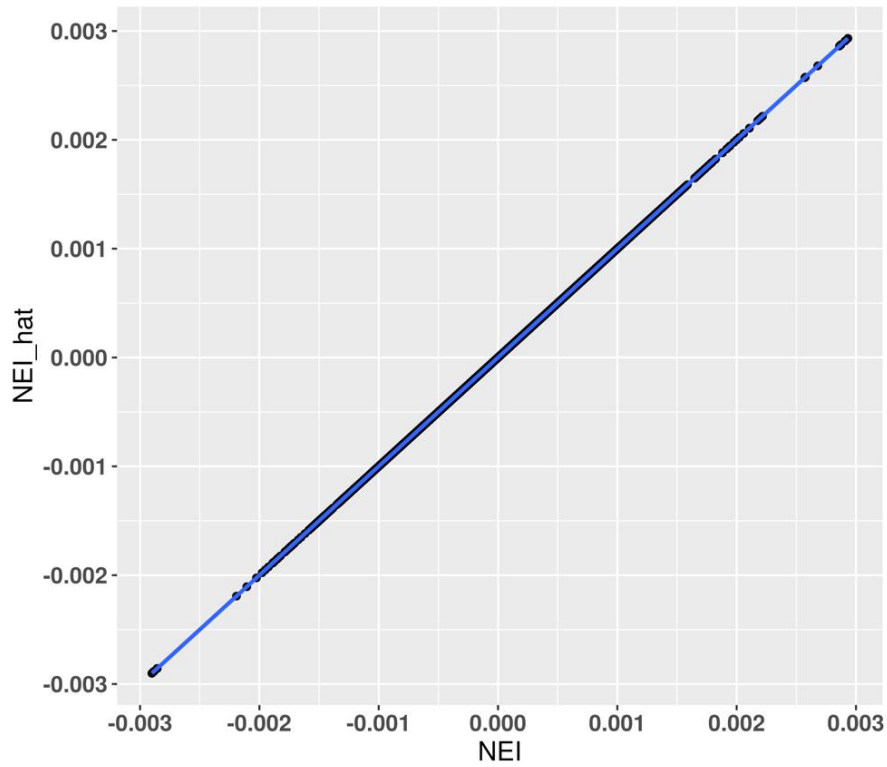
Validating genomic prediction for nitrogen efficiency index



609

610 **Figure 1.** Workflow of validating genomic prediction for nitrogen efficiency index and its
 611 composition traits. NINT1 - the N intake in primiparous cows, MTPN1 - milk true protein N in
 612 primiparous cows, MUNY1 - milk urea N yield in primiparous cows, NINT2+ - N intake in
 613 multiparous cows, MTPN2+ - milk true protein N in multiparous cows, and MUNY2+ - milk
 614 urea N yield in multiparous cows.

Validating genomic prediction for nitrogen efficiency index



615

616 **Figure 2.** Pearson correlation between SNPs effect ($n = 566,294$) is estimated directly from the
617 nitrogen efficiency index (NEI) and indirectly from its composition traits (NEI-hat). Its
618 composition traits included N intake, milk true protein N, milk urea N yield in primiparous
619 cows, and N intake, milk true protein N, milk urea N yield in multiparous cows (six traits in
620 total).