

Novel SNPs and haplotypes identified in the *CD4* gene and their influence on deregressed MACE EBV indexes of milk-related traits in Simmental breed

Research Article

Cite this article: Napolitano F, Grandoni F, De Matteis G, Degano L, Vicario D and Buttazzoni L (2021). Novel SNPs and haplotypes identified in the *CD4* gene and their influence on deregressed MACE EBV indexes of milk-related traits in Simmental breed. *Journal of Dairy Research* **88**, 368–373. <https://doi.org/10.1017/S0022029921000832>

Received: 21 June 2021
Revised: 4 August 2021
Accepted: 23 August 2021
First published online: 13 December 2021

Keywords:

Bovine; candidate gene; haplotype effect; Simmental breed

Author for correspondence:

Francesco Napolitano,
Email: francesco.napolitano@crea.gov.it

Francesco Napolitano¹, Francesco Grandoni¹, Giovanna De Matteis¹,
Lorenzo Degano², Daniele Vicario² and Luca Buttazzoni¹

¹Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) – Centro di ricerca Zootecnia e Acquacoltura, 00015 Monterotondo, Rome, Italy and ²Associazione Nazionale Allevatori Bovini di razza Pezzata Rossa Italiana, 33100 Udine, Italy

Abstract

Cluster of differentiation 4 (*CD4*) is the accessory protein non-covalently bound to the T cell receptor that recognizes an invariant region of MHC class II on antigen presenting cells. Its cytoplasmic tail, physically associated with a protein tyrosine kinase, is important in the activation of helper/inducer T lymphocytes. In *Bos taurus*, *CD4* gene is located on chromosome 5 from which two isoforms are transcribed, with a different number of amino acids due to splicing of exon 7 and variation in the reading frame. The aim of this study was to investigate the sequence of the entire *CD4* gene in Simmental sires to evaluate the effects of genomic variants on the indexes of the bulls for milk, fat and protein yields, as well as somatic cell score. The associations among genomic variants and indexes were analysed using the Allele and GLM procedures of SAS 9.4. The analysis indicated that only four of the thirty-one identified SNPs influenced the considered traits. Identified variants insist on coding zones and intronic sequences, where we revealed the presence of sites for transcription factors. To evaluate the existence of haplotypic effects, combinations among the four genomic variants (SNP 3, SNP 8, SNP 11 and SNP 19) were investigated. Six different haplotypic alleles were identified, but only four of them were frequent enough to allow for an evaluation of any haplotypic effect (at least six copies in the examined sample): Hap1, Hap2, Hap3 and Hap6. The analysis of associations between the selected haplotypes in the *CD4* gene with milk related indexes showed that bulls with Hap2 (T-A-C-C) had better indexes for milk and protein yields ($P < 0.05$), whereas the presence of the Hap1 haplotype (A-G-A-T) caused a significant decrease of the index for protein yield ($P < 0.05$). Frequencies of the two alleles Hap1 and Hap2 (9 and 36% respectively) make them of interest for their possible inclusion in breeding schemes and support the hypothesis of considering this gene as a candidate for the improvement of milk-related traits in the Simmental breed.

Cluster of differentiation 4 (*CD4*) is a single-pass transmembrane glycoprotein playing an important function in the interaction between T lymphocytes and antigen-carrying cells (König and Zhou, 2004). In *Bos taurus*, *CD4* gene is located on chromosome 5 and two isoforms X1 (GenBank Accession Number: XM_024991416) and X2 (XM_024991417) are transcribed. The coded proteins are made of 455 (XP_024847184) and 395 (XP_024847186) amino acids respectively, due to splicing of exon 7 and variation in the reading frame.

The influence of some variants in the *CD4* genomic sequence is already known in humans for their ability to increase susceptibility to HIV infection (Oyugi *et al.*, 2009) or for their association to the susceptibility and severity of rheumatoid arthritis (Hussein *et al.*, 2012). Also, in the animal field recent researches have highlighted an important role of *CD4* in immunity. In particular, Wang *et al.* (2013) found the *CD4* gene to be linked to mastitis resistance in dairy cows. He *et al.* (2011) highlighted a significant association between SNPs within *CD4* gene with mastitis indicators and milk production traits. Usman *et al.* (2016) showed a significant association between some SNPs of the *CD4* gene (between exon 2 and intron 6) and the percentage of fat in milk. Recently, Zeb *et al.* (2020) showed that four SNPs between exon 2 and exon 4 of the *CD4* gene were significantly associated with clinical mastitis incidence and with annual milk yield. Grandoni *et al.* (2020) identified in two Italian breeds (Friesian and Simmental) and their crosses two new missense mutations in the exon 5: g.103637203 G > T and g.103637191 A > C. The mutations were involved in lack or reduced labeling of anti-*CD4* monoclonal antibody clone CC8 detected by flow cytometry assay.

In the present study, the entire *CD4* gene sequences of Simmental sires were investigated in order to reevaluate the effects of genomic variants on their deregressed multiple across country

© The Author(s), 2021. Published by Cambridge University Press on behalf of Hannah Dairy Research Foundation. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.



CAMBRIDGE
UNIVERSITY PRESS

evaluation (MACE) of estimated breeding values (EBV) indexes for milk, fat and protein yields, and for somatic cell score (SCS). This index allows one to obtain the genetic evaluations of the reproducers, elaborated by the various countries, in a single international evaluation. Using the MACE model, Interbull (International Bull Evaluation Service) evaluates each trait considering the estimates of the genetic correlations between sires in the different countries. This paper documents the presence of a substantial number of polymorphisms in the *CD4* gene and their significant associations on milk and protein indexes.

Material and methods

Data used in this study came from the Italian Simmental Breeders Association (ANAPRI). The Association provided for the complete genomic sequences and the deregressed MACE EBV indexes of 34 Simmental sires of different origins (21 Austrian, 7 Swiss, 4 German and 2 Italian) and born between 1969 and 2006. The sires had an average of 6,316 daughters with a minimum of 95 daughters and a maximum of 42 657. Indexes concerning yields of milk, fat and protein as well as SCS and their genetic basis referred to the years 2013–2015, with an update happening every five weeks.

Genetic variants in the *CD4* gene

In order to investigate the presence of polymorphisms in the coding and regulatory regions of *CD4* gene, we used the complementary sequence of the genomic region of chromosome 5 in Dominette's *Bos taurus* (GeneBank, Accession Number: NC_037332). This sequence ranges between the nucleotides g.103631000...103660000 (29,000 bp) and it contains both the coding and regulatory regions of the bovine *CD4* gene.

Genome-wide sequencing of each of the 34 bulls was uploaded to Galaxy server at <https://usegalaxy.eu> (Version 2.3.4.3). Bowtie 2, tool of the Galaxy software, was used to map reads of the 34 bulls on the region of chromosome 5 including the *CD4* gene, after assessing the quality of the raw reads through Galaxy's FastQC tool.

Bioinformatic analysis

Bioinformatic analysis was performed using the MATCH™ programme (Kel *et al.*, 2003) in the TRANSFAC® Professional 10.2 <http://www.biobaseinternational.com/> (Matys *et al.*, 2006). A profile search was performed for transcription factor binding sites in the sequenced DNA fragments, focusing on the areas where SNPs had been detected.

The 'vertebrate binding' and the 'only high quality' matrix options were used in the profile selection, with cut-offs for core and matrix similarity set to 0.75 and 0.7 respectively, without any changes in other options.

Statistical analysis

On each SNP site, the χ^2 test for deviation from Hardy-Weinberg equilibrium, along with expected and observed heterozygosity, polymorphism information content (PIC) and linkage disequilibrium (LD) were calculated using the algorithms provided for by the SAS software 9.4 (ALLELE procedure).

The associations between SNPs and traits were analysed using the general linear model (GLM procedure of SAS software 9.4):

$Y_{ik} = \mu + G_i + e_{ik}$ where Y_{ik} is the phenotype, μ was the overall population mean for the trait, G_i is the fixed effect of the genotype at the SNP ($i = AA, AB, BB$) and e_{ik} is the random error. Significance of the SNP and least-squares means were determined using the Student's *t*-test in GLM procedure.

The association between haplotypes and traits were analysed by a GLM on each haplotype by inserting the number of copies as a fixed: $Y_{ik} = \mu + H_i + e_{ik}$, where Y_{ik} is the phenotype, μ is the overall population mean for the trait, H_i is the fixed effect of the haplotype [$i = 0$, individuals with no haplotype; 1, individuals with one copy only (heterozygotes); 2, individuals homozygous for that haplotype] and e_{ik} is the random error. Significance of the haplotype and least-squares means were determined using the Student's *t*-test in GLM procedure.

Results

The comparison of two already known sequences of bovine mRNA (transcript variant X1, NCBI accession: XM_024991416; transcript variant X2, NCBI accession: XM_024991417) with the DNA fragment of 29 000 bp deriving from the genomic sequence of the Dominette bovine, suggested the use of this DNA fragment as reference basis for the study of possible variants in the *CD4* gene. In fact, this genomic fragment contains both the coding and regulatory parts of the two transcript isoforms. The X1 and X2 variants differ in two respects: a different 5'UTR and the lack of exon 7 in variant X2. While the 5'UTR of the X1 variant ranges between the nucleotides g.103654819 ...103654611, the same UTR in the X2 variant ranges between g.103651793 ...103651649 of the bovine sequence (NC_037332). Furthermore, the splicing of exon 7 in the X2 variant involves a modification in the reading of the nucleotide sequence with the formation of a stop codon (TAG) in position g.103633422 and the consequent reduction of the number of amino acids in the protein isoform (395aa – in X2 – XP_024847186 as compared to the 455aa in X1 – XP_024847184).

Genome-wide sequencing analysis of 34 bulls uploaded to the Galaxy server showed that the number of raw reads per bull on average is approximately 181 million, for both forward and reverse sequencing. The quality control of these reads, produced by the FastQC report, highlights a good quality of the sequences expressed by an average Phred Score value of 38, remembering that the minimum value for a good sequence is 20 (Kwong *et al.*, 2015). Aligning the reads with the Bowtie2 tool, on the 29 000 bp genomic sequence of the *CD4* gene, showed an average of 6.2 million reads for both forward and reverse bull sequencing. Furthermore, only genomic variants with a Phred Quality Score > 40 were considered.

Diversity analysis of *CD4* SNPs

Thirty-one SNPs were identified in the described genomic sequence (Table 1) sixteen of them causing sequence variations in coding or regulatory regions of the *CD4* gene (SNPs: 1, 2, 11 ÷ 17, 21 ÷ 24, 27 and 30 ÷ 31), and the others located in regions not expected to affect gene functionality. Among those positioned in coding regions, SNPs 11 ÷ 17 (exon 2), 22 (exon 4), 23 and 24 (exon 5), produce changes in the amino acid sequence, while the other two are synonymous (21 of exon 3 and 27 of exon 6). Three SNPs were newly found in the *CD4* gene, namely: SNP 19 and 20, identified in intron 2 and SNP 30 present in the 3'-UTR regulatory region.

Table 1. Information on genetic variants of the *CD4* gene in Simmental breed.

Name	Location	Mutation type on ARS-UCD1.2 NC_037332	Mutation type on UMD-3.1.1 AC_000162	Reference	AA change
SNP 1	5'UTR	g.103654811 A > G	g.104017778 A > G	rs110248262	No
SNP 2	5'UTR	g.103654805 C > T	g.104017772 C > T	rs109536830	No
SNP 3	5'UTR ÷ Exon 1	g.103653720 T > A	g.104016687 T > A	rs211555032	No
SNP 4	5'UTR ÷ Exon 1	g.103653717 C > T	g.104016684 C > T	rs137378646	No
SNP 5	5'UTR ÷ Exon 1	g.103651620 C > T	g.104014587 C > T	rs110147215	No
SNP 6	5'UTR ÷ Exon 1	g.103649047 C > T	g.104012014 C > T	rs110810671	No
SNP 7	5'UTR ÷ Exon 1	g.103647837 C > T	g.104010804 C > T	rs110894017	No
SNP 8	5'UTR ÷ Exon 1	g.103647785 G > A	g.104010752 G > A	rs110774456	No
SNP 9	5'UTR ÷ Exon 1	g.103647727 A > G	g.104010694 A > G	rs109180561	No
SNP 10	Intron 1	g.103647395 C > T	g.104010362 C > T	rs110683263	No
SNP 11	Exon 2	g.103647315 A > C	g.104010282 A > C	rs110955838	K36Q
SNP 12	Exon 2	g.103647230 A > G	g.104010197 A > G	rs134689546	K64S
SNP 13	Exon 2	g.103647229 G > T	g.104010196 G > T	rs136742947	K64S
SNP 14	Exon 2	g.103647227 G > A	g.104010194 G > A	rs135920682	R65H
SNP 15	Exon 2	g.103647216 T > G	g.104010183 T > G	rs134722546	F69V
SNP 16	Exon 2	g.103647214 T > G	g.104010181 T > G	rs133342397	F69V
SNP 17	Exon 2	g.103647213 T > C	g.104010180 T > C	rs135440143	Y70H
SNP 18	Intron 2	g.103643593 G > C	g.104006560 G > C	rs110899347	No
SNP 19	Intron 2	g.103643031 T > C	g.104005998 T > C	Novel	No
SNP 20	Intron 2	g.103640429 T > C	g.104003396 T > C	Novel	No
SNP 21	Exon 3	g.103639414 G > A	g.104002381 G > A	rs110901299	Synonymous
SNP 22	Exon 4	g.103638641 G > A	g.104001608 G > A	rs134369392	D160N
SNP 23	Exon 5	g.103637203 G > T	g.104000170 G > T	rs109747404	Q306H
SNP 24	Exon 5	g.103637191 A > C	g.104000158 A > C	rs208394727	K310N
SNP 25	Intron 5	g.103635305 C > T	g.103998274 C > T	rs134508820	No
SNP 26	Intron 5	g.103635017 T > A	g.103997986 T > A	rs133626228	No
SNP 27	Exon 6	g.103634803 G > A	g.103997772 G > A	rs136156940	Synonymous
SNP 28	Intron 6	g.103634333 A < G	g.103997302 A < G	rs110387779	No
SNP 29	Intron 6	g.103633949 C > T	g.103996918 C > T	rs110421401	No
SNP 30	3'UTR	g.103633011 C > G	g.103996216 N > N	Novel	No
SNP 31	3'UTR	g.103632008 G > C	g.103995305 G > C	rs136340758	No

Most of the SNPs are informative markers (online Supplementary Table S1), with a PIC value over 0.20. Only genotypes at SNPs 7 and 19 were not in Hardy-Weinberg equilibrium ($\chi^2 = 21.19$ and 14.74 , respectively; $P < 0.0001$), possibly due to some sort of selection.

Correlations among SNPs are reported in online Supplementary Table S2. Sixteen SNPs (1, 9 ÷ 18, 21, 22, 25, 28 and 29), one half of the total number, were 100% in linkage disequilibrium ($LD = 1$). Out of these, one was chosen for the association analysis based on its location in the coding region of the gene and because it had already been used by other authors (Usman *et al.*, 2016; Zeb *et al.*, 2020) in association tests in dairy cattle: SNP 11 (rs110955838) in exon 2. Ten more SNPs with $LD < 0.9$: SNP 2 (rs109536830), SNP 3 (rs211555032), SNP

5 (rs110147215), SNP 7 (rs110894017), SNP 8 (rs110774456), SNP 19 (novel), SNP 20 (novel), SNP 27 (rs136156940), SNP 30 (novel), and SNP 31 (rs136340758) were processed for the estimation of their associations with phenotypes (deregressed MACE EBV indexes of milk, fat and protein yields, and SCS) as function of their genotype.

Association analysis of each SNP on milk indexes

Table 2 reports the association of the genotypes at the 11 SNP loci with the indexes on milk, fat and proteins yields, and SCS. The association analysis indicated that only four of the eleven chosen markers had an effect on some of the considered traits. In particular, SNPs: 3, 8, 11, and 19, exhibited a significant effect ($P < 0.05$)

Table 2. Association of SNPs of the *CD4* gene with the deregressed multiple across country evaluation of estimated breeding values indexes of some milk-related traits in Simmental breed.*

Name**	Genotype	N	Trait			
			Milk (kg)	Fat (kg)	Protein (kg)	SCS
			Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE
SNP 2	CC	19	-629 ± 192	-38 ± 15	-20 ± 8	92 ± 3
	CT	13	-19 ± 242	1 ± 19	-6 ± 10	91 ± 4
	TT	2	-150 ± 593	-22 ± 46	-9 ± 24	91 ± 9
SNP 3	TT	28	-277 ± 163	-24 ± 13	-7 ± 6 ^a	91 ± 3
	TA	6	-837 ± 347	-21 ± 27	-46 ± 13 ^b	93 ± 6
SNP 5	TT	28	-378 ± 169	-25 ± 13	11 ± 7	92 ± 3
	CT	6	-381 ± 359	-14 ± 27	-29 ± 14	91 ± 6
SNP 7	CC	4	-516 ± 438	-23 ± 34	-12 ± 17	99 ± 8
	CT	30	-359 ± 163	-23 ± 13	-14 ± 6	91 ± 3
SNP 8	AA	20	-208 ± 193	-24 ± 16	-3 ± 7 ^a	91 ± 3
	GA	13	-695 ± 234	-24 ± 19	-33 ± 9 ^b	93 ± 4
SNP 11	CC	26	-237 ± 168 ^A	-23 ± 13	-5 ± 6 ^a	91 ± 3
	AC	8	-820 ± 297 ^B	-24 ± 24	-44 ± 11 ^b	93 ± 5
SNP 19	TT	7	-1102 ± 298 ^b	-37 ± 25	-42 ± 6 ^b	88 ± 6
	TC	27	-183 ± 154 ^a	-19 ± 13	-7 ± 6 ^a	92 ± 3
SNP 20	TT	22	-465 ± 190	-31 ± 14	-18 ± 7	91 ± 3
	TC	12	-227 ± 251	-8 ± 19	-8 ± 10	92 ± 4
SNP 27	AA	9	-577 ± 294	-20 ± 23	-14 ± 12	94 ± 5
	GA	18	-290 ± 214	-30 ± 16	-10 ± 8	92 ± 3
	GG	7	-338 ± 334	-9 ± 26	-24 ± 13	89 ± 6
SNP 30	CC	16	-644 ± 212	-23 ± 17	-22 ± 9	94 ± 4
	CG	15	-91 ± 227	-25 ± 18	-6 ± 9	90 ± 4
	GG	3	-301 ± 490	-13 ± 39	-11 ± 20	88 ± 8
SNP 31	CC	11	-499 ± 268	-19 ± 20	-13 ± 10	93 ± 5
	CG	17	-298 ± 222	-31 ± 17	-11 ± 9	92 ± 4
	GG	6	-371 ± 362	-9 ± 28	-26 ± 14	88 ± 6

*Within each group, values with differing superscript letters mean a significant difference (a, b = $P < 0.05$) or numerical but non-significant difference (A, B = $P < 0.10$).

**Novel SNPs are marked in bold.

on protein yield index; SNP 19 exhibited a significant effect ($P < 0.05$) also on milk yield, while SNP 11 showed a numerical but non-significant difference ($P > 0.05$) on the same trait.

Bioinformatic analysis on the detected SNPs

The identification of SNPs affecting the considered indexes even if located in non-coding or regulatory regions (SNPs: 3, 8 and 19) led us to evaluate the existence of transcription factor sites near them.

Transcription factors identified in TRANFAC database were present in the regions of the *CD4* gene modified by our markers. SNP 3 T > A was located into the 'core' of the consensus sequence (ggT/AGACacctgt, g.103653722...711 on NC_037332) of binding site for transcription factors CREB (cAMP Response Element Binding Protein). The novel SNP 19 T > C was located

into the 'core' of the consensus sequence (tttGATCT/Cct, g.103643038...029 on NC_037332) for GATA-3 (GATA-binding protein 3). No transcription factors were detected in the sequence affected by the marker SNP 8.

Association analysis of haplotypes on milk indexes

As reported in Table 2, 4 out of 11 analysed SNPs (3, 8, 11 and 19) produced significant effects on milk related indexes. To evaluate the existence of haplotypic effects, combinations among the four genomic variants were investigated.

Haplotype make-up and frequencies are listed in Table 3. Six different haplotypic alleles were identified, but only four of them were frequent enough to allow for an evaluation of any haplotypic effect (at least six copies in the sample examined): Hap1, Hap2, Hap3, and Hap6.

Table 3. Information on haplotypic alleles of the *CD4* gene in Simmental breed.

Haplotype	N	SNP 3 rs211555032	SNP 8 rs110774456	SNP 11 rs110955838	SNP-19 novel	Frequency
Hap1	6	A	G	A	T	0.09
Hap2	26	T	A	C	C	0.36
Hap3	27	T	A	C	T	0.42
Hap4	2	T	G	A	T	0.03
Hap5	1	T	G	C	C	0.01
Hap6	6	T	G	C	T	0.09

Table 4. Association of the haplotypic alleles of the *CD4* gene with the deregressed multiple across country evaluation of estimated breeding values indexes of some milk traits in Simmental breed*

Trait			Milk, kg	Fat, kg	Protein, kg	SCS
Haplotype/Frequency	Copy	N	Mean ± SE	Mean ± SE	Mean ± SE	Mean ± SE
Hap1 (A-G-A-T)/9%	0	28	-277 ± 163	-24 ± 13	7 ± 6 ^a	91 ± 3
	1	6	-837 ± 347	-21 ± 28	-46 ± 13 ^b	93 ± 6
Hap2 (T-A-C-C)/36%	0	8	-901 ± 291 ^b	-31 ± 24	-36 ± 11 ^b	88 ± 5
	1	26	-211 ± 165 ^a	-20 ± 13	-7 ± 6 ^a	93 ± 3
Hap3 (T-A-C-T)/42%	0	9	-179 ± 294	-7 ± 22	-15 ± 12	93 ± 5
	1	23	-459 ± 188	-31 ± 14	-14 ± 8	92 ± 3
	2	2	-394 ± 625	-6 ± 48	-7 ± 25	74 ± 14
Hap6 (T-G-C-T)/9%	0	28	-390 ± 169	-24 ± 13	-15 ± 7	92 ± 3
	1	6	328 ± 358	-19 ± 27	-11 ± 14	91 ± 6

*Within each group, values with differing superscript differ significantly, $P < 0.05$ or greater.

The analysis of the associations between the selected haplotypes in the *CD4* gene with milk related indexes (Table 4) showed that bulls with haplotype Hap2 (T-A-C-C) had better indexes for milk and protein yields ($P < 0.05$), whereas the presence of the Hap1 haplotype (A-G-A-T) caused a significant decrease of the index for protein yield ($P < 0.05$). Frequencies of the two alleles Hap1 and Hap2 (9 and 36% respectively) make them of particular interest.

Discussion

This study documents the occurrence of several polymorphisms in the genomic sequence of the entire *CD4* gene in a sample of Simmental genomic sires. Many of these polymorphisms, as already reported by other authors (Usman *et al.*, 2016; Zeb *et al.*, 2020), were in perfect linkage disequilibrium in the surveyed 34 genomes. In the present work we investigated the possible effect of genomic variants in *CD4* on the deregressed MACE EBV indexes for SCS, an indicator of udder health status and for milk quality and quantity.

The identified variants insist on coding regions (variants of exons 2, 3, 4, 5 and 6), in regulatory regions (5' and 3' UTR) and also within intronic sequences. Attention to these last two areas of the *CD4* gene revealed the presence of sites for transcription factors.

SNP 3 (rs211555032) located into the 'core' of the consensus sequence of binding site for CREB, determined a significant and negative effect on protein yield index. CREB is a transcription

factor that regulates diverse cellular responses, including proliferation, survival, and differentiation. Its activation is required for the generation and maintenance of regulatory T cells (Wen *et al.*, 2010). Furthermore, as reported by Pegolo *et al.* (2018) the transcription factor CREB is known to regulate milk protein synthesis.

In our sample of Simmental sires the SNP 8 (rs110774456) highlights a significant and negative effect on the protein yield index. No effect was detected on the other traits, in particular for the somatic cell score.

The analysis of the genetic variants of the *CD4* gene highlights that SNP 11 (rs110955838) is in very close linkage disequilibrium with most of the variants present in the coding and regulatory areas of the gene. This is also evident from the results obtained by other authors in the analysis of the association between some SNPs of *CD4* and production traits in dairy cattle (Usman *et al.*, 2016; Zeb *et al.*, 2020). He *et al.* (2011) reported a significant association between the SNP g.13598 C > T of the *CD4* gene in intron 6, which corresponds to the SNP 29 (rs110421401) and which is in perfect linkage disequilibrium with the SNP 11, with the production traits of milk and protein yield in Chinese Holsteins, in agreement with our results on Simmental sires. In their study, the authors also highlighted a significant effect on the somatic cell score trait, which is not highlighted in our data. Other studies reported association of SNPs in *CD4* gene with fat percentage but not with SCS (Usman *et al.*, 2016), as well as Zeb *et al.* (2020) that describe a significant association with the frequency of clinical mastitis but not with

SCS. The same authors identify possible reasons for these discrepancies in the differences in the breeds of cattle, the environment conditions, and population structure.

The novel SNP 19 was located in the 'core' of the consensus sequence of binding site for the transcription factor GATA-3 and determined a significant and positive effect on milk and protein yield indexes. The zinc finger transcription factors GATA-3 was defined as a master regulator of the mammary gland (Tong and Hotamisligil, 2007): its expression is required for normal embryonic mammary development, while in the postnatal mammary gland, GATA-3 is essential for the formation of the ductal tree and for differentiation of the epithelial cells (Kouros-Mehr *et al.*, 2006; Tong and Hotamisligil, 2007; Kinoshita *et al.*, 2014). Gata-3 expression is also essential for early T cell development in the thymus and for differentiation of naïve CD4⁺ T cells into committed T helper type 2 cells (Skapenko *et al.*, 2004; Tindemans *et al.*, 2014).

The haplotypic approach allows us to gain a systemic view of the probable role of the *CD4* gene in the determinism of livestock traits. This study, within the limits of our sample, highlighted the association of some of the newly found haplotypes with milk-related deregressed MACE EBV indexes in Simmental breed. Therefore, they could be used as genetic markers in future breeding programs. For example, it should be possible to increase the frequency of haplotype Hap2 (T-A-C-C) in the population, if the selection goals included milk quality and yield improvement. Protein yield appear to be significantly depressed by Hap 1 (A-G-A-T). As it can be easily seen, the two haplotypes differ in all four SNPs contributing to their identification. This is important as an indirect confirmation of the results from SNP association analysis.

Although the role of *CD4* gene during inflammation is well documented (Oyugi *et al.*, 2009; Usman *et al.*, 2016), little is known about its possible association with production traits. This work implements knowledge on *CD4* genomic variability through a 'candidate gene' approach, highlighting three novel variants. These new SNPs, together with the many other ones previously identified by other authors, allowed us to identify new haplotypes considering the entire gene sequence. The significant association of *CD4* haplotypes with milk and protein indexes observed in this study supports the hypothesis of considering this gene as a candidate for the improvement of milk-related traits in Simmental breed.

In conclusion, we investigated the sequence of the entire *CD4* gene in Simmental sires to evaluate the effects of genomic variants on the indexes of the bulls for milk, fat and protein yields, as well as somatic cell score. Only four of thirty-one identified SNPs influenced these traits. To evaluate the existence of haplotypic effects, combinations among these four were investigated and six different haplotypic alleles were identified, of which four were frequent enough to allow for a full evaluation: Hap1, Hap2, Hap3 and Hap6. The analysis of associations between the selected haplotypes in the *CD4* gene with milk related indexes showed that bulls with Hap2 (T-A-C-C) had better indexes for milk and protein yields whereas the presence of the Hap1 haplotype (A-G-A-T) caused a significant decrease of the index for protein yield. Frequencies of the two alleles Hap1 and Hap2 (9 and 36% respectively) make them of interest for their possible inclusion in breeding schemes and support the hypothesis of considering this gene as a candidate for the improvement of milk-related traits in the Simmental breed.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0022029921000832>

References

- Grandoni F, Napolitano F, Scatà MC and De Matteis G (2020) A new polymorphic epitope of bovine CD4 antigen evidenced by flow cytometry. *Veterinary Immunology and Immunopathology* **219**, 109957.
- He Y, Chu Q, Ma P, Wang Y, Zhang Q, Sun D, Zhang Y, Yu Y and Zhang Y (2011) Association of bovine *CD4* and *STAT5b* single nucleotide polymorphisms with somatic cell scores and milk production traits in Chinese holsteins. *Journal of Dairy Research* **78**, 242–249.
- Hussein YM, El Tarhouny SA, Mohamed RH, El-Shal AS, Abul-Saoud AM and Abdo M (2012) Association of *CD4* enhancer gene polymorphisms with rheumatoid arthritis in Egyptian female patients. *Rheumatology International* **32**, 2325–2330.
- Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV and Wingender E (2003) MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Research* **31**, 3576–3579.
- Kinoshita Y, Yoshizawa K, Emoto Y, Yuki M, Yuri T, Shikata N and Tsubura A (2014) Similarity of GATA-3 expression between Rat and human mammary glands. *Journal of Toxicologic Pathology* **27**, 159–162.
- König R and Zhou W (2004) Signal transduction in T helper cells: cD4 coreceptors exert complex regulatory effects on T cell activation and function. *Current Issues in Molecular Biology* **6**, 1–15.
- Kouros-Mehr H, Slorach EM, Sternlicht MD and Werb Z (2006) GATA-3 Maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127**, 1041–1055.
- Kwong JC, McCallum N, Sintchenko V and Howden BP (2015) Whole genome sequencing in clinical and public health microbiology. *Pathology* **47**, 199–210.
- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE and Wingender E (2006) TRANSFAC[®] and its module TRANSCOMP[®]: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research* **34**, D108–D110.
- Oyugi JO, Vouriot FCM, Alimonti J, Wayne S, Luo M, Land AM, Ao Z, Yao X, Sekaly RP, Elliott LJ, Simonsen JN, Ball TB, Jaoko W, Kimani J, Plummer FA and Fowke KR (2009) A common *CD4* gene variant is associated with an increased risk of HIV-1 infection in Kenyan female commercial sex workers. *The Journal of Infectious Diseases* **199**, 1327–1334.
- Pegolo S, Mach N, Ramayo-Caldas Y, Schiavon S, Bittante G and Cecchinato A (2018) Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. *Scientific Reports* **8**, 566.
- Skapenko A, Leipe J, Niesner U, Devriendt K, Beetz R, Radbruch A, Kalden JR, Lipsky PE and Schulze-Koops H (2004) GATA-3 in human T cell helper type 2 development. *Journal of Experimental Medicine* **199**, 423–428.
- Tindemans I, Serafini N, Di Santo JP and Hendriks RW (2014) GATA-3 Function in innate and adaptive immunity. *Immunity* **41**, 191–206. <https://doi.org/10.1016/j.immuni.2014.06.006>.
- Tong Q and Hotamisligil GS (2007) Cell fate in the mammary gland. *Nature* **44**, 724–726.
- Usman T, Yu Y, Zhai L, Liu C, Wang X and Wang Y (2016) Association of *CD4* SNPs with fat percentage of Holstein cattle. *Genetics and Molecular Research* **15**, gmr.15038697.
- Wang XS, Zhang Y, He YH, Ma PP, Fan LJ, Wang YC, Zhang YI, Sun DX, Zhang SL, Wang CD, Song JZ and Yu Y (2013) Aberrant promoter methylation of the *CD4* gene in peripheral blood cells of mastitic dairy cows. *Genetics and Molecular Research* **12**, 6228–6239.
- Wen AY, Sakamoto KM and Miller LS (2010) The role of the transcription factor CREB in immune function. *The Journal of Immunology* **185**, 6413–6419.
- Zeb S, Ali N, Niaz S, Rasheed A, Khattak I, Khan NU, Wang Y and Usman T (2020) Association of SNPs in the coding regions of *CD4* gene with mastitis susceptibility and production traits in dairy cattle. *Thai Journal of Veterinary Medicine* **50**, 75–80.