RESEARCH ARTICLE

Complement component 3: characterization and association with mastitis resistance in Egyptian water buffalo and cattle

NERMIN EL-HALAWANY[1]*, ABD-EL-MONSIF A. SHAWKY[1], AHMED F. M. AL-TOHAMY[1], LAMEES HEGAZY[2], HAMDY ABDEL-SHAFY[3], MAGDY A. ABDEL-LATIF[4], YASSER A. GHAZI[5], CHRISTIANE NEUHOFF[6], DESSIE SALILEW-WONDIM[6] and KARL SCHELLANDER[6]

[1]*Department of Cell Biology, National Research Centre, 12622 Giza, Egypt*
[2]*Avisina Group, Benha, Qalubia 13776, Egypt*
[3]*Faculty of Agriculture, Department of Animal Production, Cairo University, 12613 Giza, Egypt*
[4]*Department of Buffalo Research, Animal Production Research Institute, Doki, 12126 Giza, Egypt*
[5]*Department of Animal Reproduction and A.I. National Research Centre, 12622 Giza, Egypt*
[6]*Institute of Animal Science, Animal Breeding and Husbandry Group, University of Bonn, Bonn 53115, Germany*

Abstract

Mastitis is an infectious disease of the mammary gland that leads to reduced milk production and change in milk composition. Complement component C3 plays a major role as a central molecule of the complement cascade involving in killing of microorganisms, either directly or in cooperation with phagocytic cells. C3 cDNA were isolated, from Egyptian buffalo and cattle, sequenced and characterized. The C3 cDNA sequences of buffalo and cattle consist of 5025 and 5019 bp, respectively. Buffalo and cattle C3 cDNAs share 99% of sequence identity with each other. The 4986 bp open reading frame in buffalo encodes a putative protein of 1661 amino acids—as in cattle—and includes all the functional domains. Further, analysis of the C3 cDNA sequences detected six novel single-nucleotide polymorphisms (SNPs) in buffalo and three novel SNPs in cattle. The association analysis of the detected SNPs with milk somatic cell score as an indicator of mastitis revealed that the most significant association in buffalo was found in the C>A substitution (ss: 1752816097) in exon 27, whereas in cattle it was in the C>T substitution (ss: 1752816085) in exon 12. Our findings provide preliminary information about the contribution of C3 polymorphisms to mastitis resistance in buffalo and cattle.

[El-Halawany N., Abd-El-Morif S. A., Al-Tohamy A. F. M., Hegazy L., Abdel-Shafy H., Abdel-Latif M. A., Ghazi Y. A., Neuhoff C., Salilew-Wondim D. and Schellander K. 2017 Complement component 3: characterization and association with mastitis resistance in Egyptian water buffalo and cattle. *J. Genet.* **96**, 65–73]

Introduction

Mastitis is defined as an inflammatory reaction of the parenchyma of the mammary gland caused due to infection, traumatic or toxic. It has serious effects on the mammary gland like reducing milk production and altering milk composition (Oviedo-Boyso *et al.* 2007). Although, buffalo has been traditionally considered less susceptible to mastitis than cattle, the major problem in buffalo productivity is a high mortality rate of the calves in the first 3 months of life (Akhtar and Ali 1994). This could be due to the effects of maternal mastitis; the quarterwise prevalence of intramammary infection in buffalo reaches up to 79% during the first 3 months after calving (Moroni *et al.* 2006). As mastitis is an infectious disease, genes coding for immune factors, that detect and eliminate pathogens, are strong potential candidates for predicting mastitis resistance. Several studies have been conducted to identify various aspects of genetic variability in the immune mechanisms underlying mastitis resistance, including genes such as CD14, CD18, lactoferrin, lysozyme, class I and class II genes of MHC, bovine leukocyte antigen (BoLA) and toll-like receptor 4 (reviewed in Detilleux 2002; Rupp and Boichard 2003; Ibeagha-Awemu *et al.* 2008). Further, quantitative trait loci (QTLs) related to mastitis resistance and/or somatic cell count (SCC) have been identified on most chromosomes of the bovine genome, with the highest densities in BTA3 and BTA14 (reviewed in Ogorevc *et al.* 2009). Recently, more emphasis has been placed on genomewide association studies (GWAS) to

*For correspondence. E-mail: nelhalawany2001@hotmail.com, nk.elhalawany@nrc.sci.eg.

Keywords. complement component C3; genetic variants; somatic cell score; association analysis; buffalo; cattle.

identify genomic markers associated with incidence of and resistance to mastitis (Cole *et al.* 2011; Meredith *et al.* 2012; Wijga *et al.* 2012; Abdel-Shafy *et al.* 2014; Nani *et al.* 2015). Although, GWAS offer advantages in discovering genetic variants that influence disease, they also have important limitations, including the potential for false-positive and false-negative results, the lack of genetic variants of relevance in the genotyping platform, and lack of variation in SNP in the population under study (Pearson and Manolio 2008). Therefore, the constant search for candidate genes related to important economic traits alongside the study of their functional characterization should continue.

The complement system serves as the first line of defence against invading pathogens (Walport 2001; Ricklin *et al.* 2010); it consists of a complex group of plasma glycoprotein components, regulatory proteins and receptors. In particular, the complement fragment of C3 plays a major role as a central molecule of the complement cascade involving in killing of microorganisms, either directly or in cooperation with phagocytic cells (Frank and Fries 1991; Tomlinson 1993; Wanger and Frank 2009). The capacity of bovine milk to deposit complement C3 on mastitis caused by *Streptococcus agalactiae* has been reported in several studies (Rainard and Poutrel 1995; Barrio *et al.* 2003; Rainard 2003). Moreover, microarray and quantitative real-time polymerase chain reaction (qRT-PCR) analysis have shown that C3 expression level was significantly higher in the alveolar mammary tissue of quarters showing clinical signs of mastitis than in healthy, noninfected quarters (Swanson *et al.* 2009). Hence, the complement *C3* gene could be a potential candidate for mastitis resistance.

Bovine *C3* gene located on chromosome 7, its sequence and structure have been defined previously by Fredslund *et al*. (2006) and Firth *et al.* (2006). A total of 588 single-nucleotide polymorphisms (SNPs) within the bovine C3 coding region were recently reported in the GenBank database (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=280677). However, to the best of our knowledge, no association studies have been conducted between these variants and any traits. In buffalo, a predicted C3 sequence has been recently annotated in the NCBI database (accession no. XM_006045102). However, no polymorphisms in the buffalo *C3* gene have yet been reported in the NCBI SNP database.

The aim of the present study was to explore the innate resistance of Egyptian buffalo and cattle to mastitis using complement C3 as a candidate gene. We present sequence analysis of the C3 cDNA in Egyptian buffalo and cattle, information on its genetic variants and its association with milk SCS as indicator of mastitis in both animals.

Materials and methods

Animals and phenotype data

This study was carried out on both buffalo and cattle. The sampled animals were divided into two groups. In group one animals, we aimed to identify the *C3* gene sequence; liver samples were collected from five unrelated Egyptian water buffaloes and five unrelated Egyptian Baladi cattle from a slaughter house. The second group was selected for SNP genotyping and association analysis, and it comprised of 120 unrelated multiparous Egyptian water buffaloes and 120 unrelated multiparous cattle (three-generation F_2 Baladi-Frisian crossbred). These animals reared at El-Gemiza Station in El-Gharbia that belongs to the Animal Production Research Institute in Giza, Egypt. We decided to sample the Baladi-Frisian crossbred population of cattle instead of the Baladi population based on the availability of reliable information in terms of phenotypic data and management. The data on SCC were collected from the records available at the farm, which covered a period of 10 years (2004–2013).

Blood sampling, DNA and RNA isolation and cDNA synthesis

Blood samples were collected aseptically by jugular vein puncture. Genomic DNA was isolated from blood using the phenol–chloroform extraction and ethanol precipitation method (Sambrook *et al.* 1989). Liver samples were collected from the slaughter house and frozen immediately in liquid nitrogen. Total RNA was extracted from the liver tissues of five unrelated buffaloes and from five unrelated cattle using Trizol reagent (Invitrogen, Carlsbad, USA). The isolated RNA was then reverse transcribed into cDNA using a RevertAid First Strand cDNA Synthesis kit (Thermo Scientific, Waltham, USA) according to the manufacturer's instructions.

Primer design and PCR amplifications

A total of 12 specific primer pairs were designed using the online oligonucleotide design tool Primer3 (table 1) covering the C3 coding region according to the cattle C3 sequence (accession no. NM_001040469.2) to amplify overlapping fragments of 400–600 bp length. Polymerase chain reaction (PCR) was performed in a final volume of 20 μL containing 25 ng of cDNA, 10 ng of each primer, 10 mM dNTPs in a 1× PCR buffer and 0.5 U of *Taq* polymerase. The PCR conditions were carried out as follows: 95°C for 5 min, 32 cycles of 94°C for 30 s, 56–63°C (based on the primers annealing temperatures) for 30 s and 72°C for 40 s, followed by a further 5 min extension at 72°C. The PCR products were then purified using an ExoSAP-IT kit (Affymetrix UK ltd, Wycombe La High Wycombe, UK) and sequenced using a CEQ8000 sequencer system (Beckman Coulter, Brea, USA).

Gene sequence analysis

Gene sequence analysis was performed to identify the most important areas or domains of the gene and protein. The amino acid sequences of the protein were predicted using an EXPASY tool (http://au.expasy.org/translate/). The open reading frame (ORF) and signal peptide sequence of complement C3 were predicted using the ORF-Finder

Table 1. Primers used to amplify C3 cDNA of buffalo and cattle.

	Forward primer	Reverse primer	Product size
1	CTCTATCCCTCCTTCCCTCCACT	GTAGATGGTCTTGTCCGTCTGGAT	448
2	CAGCAAAGAGTTAAAATCCGACAAG	CATCTGGGTCATCAATGTAGTAGAA	470
3	CTGAGTTTGAGGTGAAGGAATATGT	GTTTGTTCTGTGTGTTGATGCTTAG	588
4	GACCTCATGGTGTACGTGACAAA	TGGCATTAATCAGTGTGTTATAGGC	578
5	AAGATCCGCTACTACACCTACATGA	CTTTATCCATCCTCTTCTCCATGA	592
6	GATGCAGGACTCACTCTCAAGAC	GAGACGCAGGTCGATAAAGAAAT	588
7	GTCAGACAAGAAAGGGATCTGTG	GTGTCTGGGACTTGATCACTAAGAT	495
8	AGAAGGCGTAAGAGTCAACAAAACT	CCAATCATTTCTTGGTGTATCACAG	594
9	GAAGCAGAAGCCTGATGGAATCT	GATCAGGAACATCCTTCTGGTATT	536
10	AGCAGAGATACTATGGAGGTGGTTA	CAGGTAGATGATGAGTGTGTTCTTG	569
11	ATGATCCTTGACATCTGTACCAAGT	ATCTTCCAGCTTCTTCTGGATAAGT	516
12	AGCAAACTCTGCCACAAAGATAC	GACAGTGTGTATGAGAGGGTGTATG	526

(http://www.ncbi.nlm.nih.gov/orffinder) and the SignalP (http://au.expasy.org/tools) computational tools, respectively. The biophysical characteristics of the predicted proteins were estimated using the ProtParam computational tool (http://au.expasy.org/tools), and the conserved protein domains of the C3 were analysed using the Smart program (http://smart.embl-heidelberg.de/). The DiANNA 1.1 web server was used to determine cysteine state and predict the disulphide bond partner (http://clavius.bc.edu/%7Eclotelab/DiANNA/). Glycosylation sites were predicted by the NetNGlyc 1.0 server: (http://www.cbs.dtu.dk/services/NetNGlyc/). The SIFT program (http://sift.bii.a-star.edu.sg) was used to determine the potential deleterious effect of amino acid changes on protein function.

Mutation detection and genotyping

To detect any nucleotide variations between different individual animals, the C3 cDNA sequences of five unrelated animals from each species were compared among each other using the NOVOSNP software package (http://www.molgen.ua.ac.be/bioinfo/novosnp/index.html). The detected SNPs were genotyped for the genomic DNA of 120 buffaloes and 120 cattle. Genotyping was done using Sequenom iPLEX Gold on the Sequenom platform offered by the Neogen genotyping service (www.neogen.com, USA).

Molecular modelling

The 3D structures of wild-type and mutated buffalo complement C3 proteins were obtained through homology modelling using the Chimera interface to Modeller (Šali and Blundell 1993; Pettersen *et al.* 2004). The X-ray crystal structure of the bovine complement component 3 (2B39) (Fredslund *et al.* 2006) was used as a template. The models were minimized under the CHARMM27 force field for proteins using the conjugate gradient algorithm implemented in NAMD 2.9 (MacKerell *et al.* 1998, 2004, Phillips *et al.* 2005). The 3D structures were solvated and neutralized using the solvation and ionization tool in VMD (Humphrey *et al.* 1996). Inspections of the molecular interactions were performed using Discovery Studio Visualizer (Dassault Systèmes 2015).

Statistical analysis

As the distribution of SCC is usually skewed, a logarithmic transformation function was adopted to obtain a distribution that is closer to normal (Ali and Shook 1980), which creates a new variable called somatic cell score (SCS) according to the formula:

$$\mathrm{SCS} = \log_2\left(\frac{\mathrm{SCC}}{100}\right) + 3.$$

Allele and genotype frequencies were calculated by direct counting. Allelic substitution effect on milk SCS was estimated according to the method suggested by Falconer and Mackay (1996):

$$\alpha = a + d(q - p),$$

where a, additive effect; d, dominance effect; q, frequencies of major allele; p, frequencies of minor allele. Additive effect was estimated as the difference between the two homozygous mean divided by 2, ($A = (\mu A11 - \mu A22)/2$), whereas the dominance effect was estimated as the deviation of the heterozygote from the mean of the two homozygotes:

$$D = (\mu A11 + \mu A22)/2 - \mu A12.$$

The detected SNPs were tested for deviation from the Hardy–Weinberg equilibrium (HWE) using GenAlEx 6.5 (Peakall and Smouse 2012).

Association of the C3 genotypes with SCS was done using the general linear model (GLM) option in SAS, ver. 9.4 (SAS 2014, SAS Institute, Cary, USA). According to the available data, the following model was applied for buffalo data analysis:

$$\mathrm{Yijkl} = \mu + \mathrm{HYSi} + \mathrm{bj1\ (DIM)} + \mathrm{bj2\ [exp(-0.05 * DIM)]} + \mathrm{gk} + \varepsilon\mathrm{ijkl},$$

where Yijkl is lactation average SCS; μ is the overall mean of observations; HYSi is the fixed combined effect of herd (h),

year (y) and season (s) of calving, where seasons were defined as calendar quarters: January–March, April–June, July–September and October–December; bj1 and bj2 are two regression coefficients associated with the fixed lactation function; DIM is days in milk (Wilmink 1987); gk is the random effect of SNP genotypes and εijkl is the residual error. Post-hoc differences among genotype classes were tested for significance using a Tukey–Kramer test as implemented in SAS (SAS 2014, SAS Institute).

In cattle, we used the following model for analysis based on the data available for the cattle populations at the station:

$$\text{Yijkl} = \mu + \text{HYSi} + \text{bj (MY)} + \text{gk} + \varepsilon\text{ijkl},$$

where Yijkl is lactation average SCS; μ is the overall mean of observations; HYSi is the fixed combined effect of herd (h), year (y) and season (s) of calving, where seasons were defined as calendar quarters: January–March, April–June, July–September and October–December; bj (MY) is the regression coefficient associated with the fixed effect of milk yield (Jamrozik and Schaeffer 2010); gk is the random effect of SNP genotypes and εijkl is the residual error. Post-hoc differences among genotype classes were tested for significance using a Tukey–Kramer test implemented in SAS (SAS 2014).

Results

C3 cDNA sequence analysis and characteristics of the derived amino acid sequence of buffalo and cattle

Buffalo and cattle *C3* gene sequences were assembled with the assistance of the overlapping regions of the PCR products, resulting in 5025 and 5019 bp cDNA contigs in buffalo and cattle, respectively. The resulting contigs were submitted to GenBank (accession nos: KP176673.1 and KP221292.1, respectively). Basic local alignment search tool (BLAST) analysis revealed that the C3 cDNA in buffalo and cattle shared 99% of nucleotide identity with each other. The predicted C3 ORF in buffalo is 4986 bp in length, as in cattle. Alignment of the C3 cDNA of buffalo and cattle with other animals' sequences in GenBank revealed that the C3 cDNA of both animals showed identities of 99, 99, 87 and 86% with those predicted for buffalo (XM_006045102.1), cattle (XM_010806886.1), pig (NM_214009.1) and Arabian camel (XM_010998272.1), respectively, whereas alignment with human C3 cDNA (NM_009778.1) revealed identities of 82% for buffalo and 81% for cattle.

The C3 amino acid (aa) sequences of buffalo and cattle were predicted based on the nucleotide sequence. Buffalo C3 protein shares 99% of amino acid identity with that of cattle and includes all the known functional C3 domains. The ORF of 4986 bp in both animals encode a putative protein of 1661 amino acid residues. Further analysis of the deduced polypeptide using the SignalP 3.0 server (http://www.cbs.dtu.dk/services/SignalP/) predicted a secretion leader sequence from amino acids 1–22 in both animals, which is similar to that in human C3 (de Bruijn and Fey 1985).

Using bioinformatics tools and published information, we identified the conserved domains of buffalo C3 protein. The predicted conserved domains of buffalo C3 are 100% match with those of cattle, as described by Fredslund *et al.* (2006). The predicted conserved domains are presented in table 2.

C3 protein consists of a β chain of 643 amino acids (23–665) and an α chain of 992 amino acids (670–1661). Both chains are linked through the tetraarginine linker that is seen in the buffalo C3 protein at the same position as in cattle (666–669).

In buffalo, the complement C3 protein possesses 27 cysteines, like that of cattle and humans. The positions of cysteine residuals that create S–S bonds between C3 fragments are similar in cattle and buffalo. The cysteine residues in buffalo were presumed to form 13 disulphide bonds, as in cattle and humans, according to calculations performed with the DiANNA 1.1 software. Disulphide bonds are important for stabilizing the structure of this protein (DiScipio and Hugli 1989).

Two N-glycosylation sites were predicted at aa positions 938 and 1649 in both buffalo and cattle protein, whereas human C3 protein was found to have two N-glycosylation sites at positions 85 and 939 (Welinder and Svendsen 1986). N-linked glycosylation is vital for the protein molecule since it supports the existence of the molecule either in membranous or soluble form.

Seven sites for O-linked glycosylation were predicted in C3 buffalo at aa positions 647, 652, 665, 673, 737, 884 and 1365, whereas in cattle, eight sites were predicted at aa positions 288, 647, 652, 665, 670, 737, 747 and 1365. Glycosylation is needed when hydrophilic clusters of carbohydrates alter the polarity and solubility of protein or protein folding.

Polymorphism detection

Polymorphisms in the complement *C3* gene of both Egyptian water buffalo and Baladi cattle were identified by comparing cDNA sequences of five unrelated animals from each species. The detected nucleotide variants of cattle C3 were compared to the NCBI SNP database (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=280677) to check their novelty. However, no reported SNPs in buffalo C3 were found. Six novel SNPs were detected in buffalo (table 3). Two of them were synonymous: SNP1B (C>T, ss: 1752816095), which is

Table 2. Predicted conserved domain in Egyptian buffalo C3 protein.

Chain	Domain	aa position in buffalo and cattle C3
β-Chain	A2M_N_2	454–603
α-Chain	A2M_C	768–865
	A2M_comp	1054–1285
	ANATO	691–726
	TED	1008–1012
	CUB	932–985 and 1287–1305
	C345c	1531–1643

Table 3. Genetic variants detected in the buffalo C3 coding region by comparing cDNA sequences of five unrelated Egyptian buffalo population.

SNP	Acc. no.	SNP pos.	Exon	Nucleotide change	aa Change	aa Position	Domain
SNP1B	ss: 1752816095	171871	12	C>T	His>His	455	A2M_N_2
SNP2B	ss: 1752816096	176258	17	A>T	Ser>Cis	686	ANATO
SNP3B	ss: 1752816097	188824	27	C>A	Ala>Ala	1159	A2M_Comp
SNP4B	ss: 1752816098	194006	29	T>C	Met>Thr	1248	A2M_Comp
SNP5B	ss: 1752816099	195656	30	A>G	Lys>Arg	1307	CUB
SNP6B	ss: 1752816100	200880	41	T>A	Trp>Arg	1629	C345c

SNP position based on water buffalo assembly 2.0 UMD + CASPUR (Scafold, NW_005783695.1); aa, amino acid.

located within the conserved domain A2M_N_2 (exon 12) and SNP3B (C>A, ss: 1752816097), which lies in the A2M_Comp domain (exon 27). The remaining four SNPs are nonsynonymous, namely: SNP2B (A>T, ss: 1752816096), which changes Ser>Cis in the anaphylatoxin (ANATO) domain (exon 17); SNP4B (T>C, ss: 1752816098), which changes Met>Thr in the A2M_Comp domain (exon 29); SNP5B (A>G, ss: 1752816099), which changes Lys> arginine (Arg) in the CUB domain (exon 30) and SNP6B (T>A, ss: 1752816100), which changes Trp>Arg in the C345c domain (exon 41).

Similarly, comparison of C3 cDNA from five different Egyptian cattle revealed three novel SNPs (table 4). The first one (SNP1C, C>T, ss: 1752816085) is synonymous and exists inside the A2M_N_2 domain (exon 12). The remaining two are nonsynonymous: The first is SNP2C (A>T, ss: 1752816086), which occurs within the ANATO domain (exon 17) and changes Ser>Cis. The other is SNP3C (T>A, ss: 1752816087), which is located in the C345c domain (exon 41) and changes Trp>Arg. Interestingly, the three SNPs that were detected in cattle were also found in buffalo; their corresponding SNPs in buffalo are SNP1B, SNP2B and SNP6B, respectively.

Effect of SNP6B (Trp1629Arg) on the protein structure of C3 in buffalo

Functional analysis of the detected SNPs using the SIFT program (http://sift.bii.a-star.edu.sg) predicted a deleterious effect on C3 protein due to SNP6B and its corresponding mutation in cattle SNP3C. This mutation led to substitution of Trp1629 to Arg in the C345c domain. Based on this prediction, we generated structural models to evaluate the effect of mutation (Trp1629Arg) on the protein structure (figure 1).

Molecular modelling demonstrated that the aa tryptophan (Trp1629) in C345c maintains various favourable, nonbonding interactions with surrounding residues, such as hydrogen bonding interactions with His1599, electrostatic pi–cation interactions with Lys1544 and pi–alkyl interactions with amino acids Lys1544, Leu1601 and Leu1647. SNP6B, which changes tryptophan (Trp1629) to arginine (Arg), maintains the hydrogen bonding interaction with the backbone of His1599 and hydrophobic alkyl interactions with Leu1647; however, electrostatic and hydrophobic pi-interactions are lost. Moreover, replacement of W1629 with Arg leads to unfavourable positive–positive charge repulsive interactions with Lys1544.

Statistical analysis

Alleles, genotype frequencies and allele substitution effects in both buffalo and cattle are shown in table 5. SNPs with minor allele frequencies less than 0.05 in the populations (SNP4B, SNP5B, SNP6B, SNP2C and SNP3C) were excluded from the association analysis.

In cattle, none of the genotyped SNPs significantly deviated from the HW proportion ($P < 0.001$). In buffalo, three SNPs (SNP2B, SNP4B and SNP5B) showed a significant deviation using the same threshold. It was expected that some SNPs could deviate from the HW proportion in buffalo because these animals are under selection.

Association analysis was performed to evaluate the effect of C3 genetic variants of buffalo and cattle on the milk SCS as a mastitis indicator. In buffalo, SNP1B and SNP3B showed significant effect on SCS (P values: 0.036 and 0.0042, respectively). The genotype *CC* of SNP1B significantly

Table 4. Genetic variants detected in the cattle C3 coding region by comparing cDNA sequences of five unrelated Egyptian cattle.

SNP	Acc. no.	SNP pos.	Exon	Nucleotide change	aa Change	aa Position	Domain
SNP1C	ss: 1752816085	18996575	12	C>T	His>His	455	A2M_N_2
SNP2C	ss: 1752816086	19000847	17	A>T	Ser>Cis	686	ANATO
SNP3C	ss: 1752816087	19025430	41	T>A	Trp>Arg	1629	C345c

SNPs position based on cattle chromosome 7; aa, amino acid.

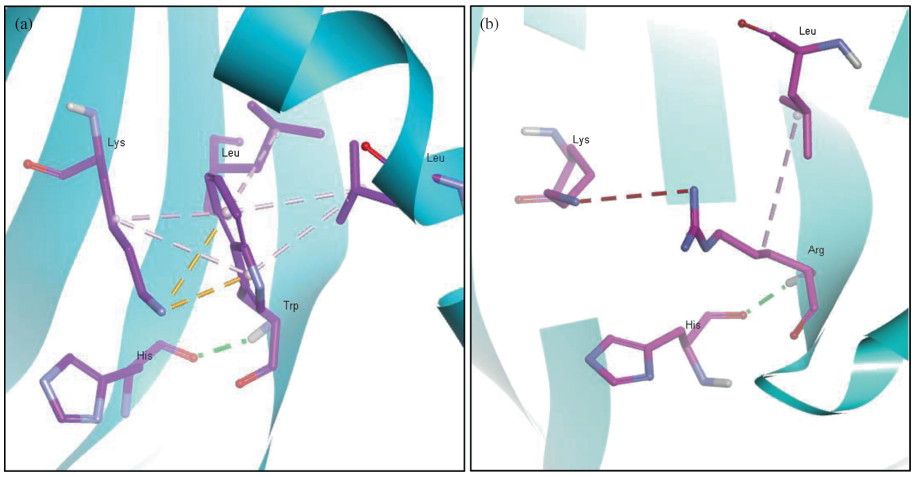

Figure 1. W1629R results in an altered electrostatic and hydrophobic interactions. (a) Wild-type complement 3, Trp1629 engages in electrostatic interactions with K1544 (orange), alkyl-pi interactions (light purple), hydrogen bonding (green). (b) Mutant W1629R, Arg1629 engages in hydrogen bonding (green) and less alkyl nonbonding interaction (light purple), positive–positive repulsive interactions (red). Interactions are demonstrated by broken line. Images generated by Discovery Studio 4.1 Visualizer.

decreased the SCS level compared to genotype *CT*. The genotype *TT* was not detected, probably due to small sample size. The substitution effect of allele C reduced SCS by 0.16. Similarly, buffaloes with the homozygous genotype *CC* of SNP3B have significantly lower SCS value than those with genotype *AA*. Allele C reduced SCS by 0.36 (table 5).

In cattle, the association analysis revealed that only SNP1C affects the SCS significantly. The genotype *CC* showed an SCS value significantly lower than the genotypes *TT* and *CT* ($P = 0.0006$). The substitution effect of allele C reduced SCS by 0.40.

Discussion

Analysis of *C3* gene sequence revealed a high degree of similarity between buffalo and cattle. These animals share 99% nucleotide and amino acid sequence identity with each other. Buffalo C3 protein includes all the known functional C3 domains, such as the TED domain, which is responsible for covalent attachment of the C3b fragment to its target (de Bruijn and Fey 1985), and the CUB domain, which plays a critical role together with the TED domain in host protection (Janssen *et al.* 2005). In addition, the thioester-associated catalytic histidine, which functions in the formation of a covalent bond between C3b and its target molecules (Law and Dodds 1997), and the neighbouring proline residues are thought to be important in the formation, stability and function of this region (Isaac and Isenman 1992).

By comparing the nucleotide sequences of five unrelated animals, we detected six novel SNPs in buffalo and three novel SNPs in cattle. Some SNPs showed very low frequencies in the genotyped animals, which could be due to small sample size.

Table 5. Associations analysis of C3 variants in Egyptian buffalo and cattle with milk SCS.

SNP	Minor allele frequency	Allele substitution effect	*P* value	Genotype (F)	SCS LSM ± SE
SNP1B	T (0.05)	−0.16	0.036	CC (0.909)	3.34 ± 0.11^{a}
				TC (0.091)	3.33 ± 0.24^{b}
SNP2B	T (0.48)	1.42	0.3783	AA (0.035)	2.36 ± 1.14
				AT (0.965)	3.36 ± 0.11
SNP3B	A (0.26)	−0.36	0.0042	AA (0.087)	2.88 ± 0.85^{ab}
				AC (0.348)	3.76 ± 0.17^{a}
				CC (0.565)	3.27 ± 0.11^{b}
SNP1C	T (0.17)	−0.40	0.0006	CC (0.691)	3.42 ± 0.10^{b}
				CT (0.284)	3.97 ± 1.15^{a}
				TT (0.025)	4.67 ± 0.35^{a}

LSM, least square means; SE, standard error. Allele substitution effect $(\alpha) = a + d(q - p)$, where a is the additive effect, d is the dominance effect, q and p are the frequencies of major and minor alleles, respectively. In the same SNP, LSM with different letters are significantly different according to a Tukey–Kramer.

Because of the importance of the complement system in defence against invading pathogens—either directly or in cooperation with phagocytic cells—the association of C3 variants with mastitis resistance using SCS as an indicator was studied in buffalo and cattle. The results showed that two SNPs in buffalo and one SNP in cattle are significantly associated with SCS. Interestingly, SNP1B in buffalo and its correspondent in cattle (SNP1C) were found to be significantly associated with SCS, which supports the contribution of this SNP to mastitis resistance. Although, the polymorphic sites that showed significant associations with milk SCS in both buffalo and cattle are silent base substitutions, these lie in important conserved domains, such as the α2-macroglobulin family N-terminal region (A2M_N_2) and the A2-macroglubin complement component (A2M_comp). The latter includes the complement components region of the α2-macroglobulin family and the ANATO domain, which is released upon activation of C3 and involved in the inflammatory response (Morgan *et al.* 2005). Recently, Kimchi-Sarfaty *et al.* (2007) and Komar (2007) reported that protein products with the same amino acid sequence, but different nucleotide sequences have different structural and functional properties. In this study, SNP1B in buffalo and its corresponding SNP1C in cattle (His>His) were synonymous, but decreased the codon frequency from 30.5 (per thousand) for GCC to 14.3 for GCA. SNP3B (Ala>Ala) changed the codon frequency from 15.5 for CAC to 9.4 for CAT according to the data available on http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9913. The changes in codon frequencies were proposed to affect the timing of the post-translational folding of proteins (Kimchi-Sarfaty *et al.* 2007). This suggestion could explain the significant association of the C3 polymorphisms with milk SCS in this study. Another explanation is that those SNPs are not causative polymorphisms, but they could be linked to other polymorphic sites or crucial QTL regions.

Using the SIFT program, we predicted that the W1629R (SNP6B) mutation, which occurs in the C345c domain, would have a deleterious effect on C3 protein. Based on this prediction, we generated structural models and performed energy minimization calculations on the wild type and mutated buffalo complement C3 proteins. This computational strategy provided a qualitative understanding of differences in the molecular interactions of amino acid residues in the C345c domain with either Trp or Arg (figure 1). Replacement of W1629 with Arg leads to unfavourable positive–positive charge repulsive interactions with Lys1544 and loss of nonbonding interactions (electrostatic and hydrophobic pi-interactions), which affect the intramolecular stability of the protein and may cause inappropriate folds of the C345c domain, thus reducing its activity.

Although, the function of the C345c domain is still unclear, it is much conserved in C3, C4 and C5. The mammalian C345c domain of C3b is involved in the interaction with the von Willebrand type A domain of Bb in the alternative pathway of C3 convertase, C3bBb (Torreira *et al.* 2009). Further, a homozygous mutation in C345c of C3 was detected in a patient with recurrent and severe infections caused by *Streptococcus pneumoniae* and associated IgA deficiency (Santos-Valente *et al.* 2013), whereas mutations in C4 and mannan-binding lectin serine peptidase 2 residues at the C345c–CCP (complement control protein) interface inhibit intermolecular interaction and C4 cleavage (Kidmose *et al.* 2012). Moreover, it has been shown that the C345c domain in C5 is important for binding to C6/7 and for the formation of nonreversible membrane attack complex (MAC) (Bramham *et al.* 2005). However, in the present study, the mutation of Trp1629 to Arg in the C345c domain of buffalo (SNP6B) and cattle (SNP3C) showed a very low minor allele frequency (MAF) (0.02) in the studied populations. This may be due to the small sample size.

The association of the *C3* gene with inflammation and pathogen-specific immune responses has been indicated previously in humans by the poor response to vaccinations in individuals carrying a deficient *C3* gene (Hazlewood *et al.* 1992; Botto *et al.* 2009; Goldberg *et al.* 2011). Moreover, polymorphisms on C3 have been seen to be associated with increased susceptibility to infections, including bacterial infections (e.g. pneumonia, bacteraemia, meningitis and osteomyelitis), which are caused by encapsulated pyogenic bacteria (reviewed in Skattum *et al.* 2011; Ricklin and Lambris 2013), as well as viral-like systemic infections (e.g. lupus erythematosus (SLE) and age-related macular infections) (Yates *et al.* 2007; Miyagawa *et al.* 2008). In animals, a deletion of cytosine at aa position 2136 was found to be responsible for hereditary canine C3 deficiency and increased susceptibility to bacterial infections (Ameratunga *et al.* 1998). In pigs, mutations in C3 were reported to be associated with complement activity and natural resistance to microorganisms (Wimmers *et al.* 2003). The previous findings, together with the results obtained in our study, suggest the possible contribution of C3 variants to resistance/susceptibility to mastitis in buffalo and cattle.

Conclusion

In the present study, the entire coding region of the complement *C3* gene of Egyptian buffalo and cattle was isolated and characterized. Buffalo and cattle C3 cDNA share a high identity of nucleotide sequence with each other. Six SNPs in buffalo and three SNPs in cattle were newly reported. The mutation of Trp1629 to Arg in the C345c domain of buffalo and cattle was predicted to have a deleterious effect on the C3 protein. Two SNPs in buffalo and one SNP in cattle significantly affect milk SCS levels. This is the first association study between C3 variants and milk SCS. Our findings provide preliminary information about the contribution of *C3* gene polymorphisms in mastitis resistance/susceptibility in buffalo and cattle. However, further large-scale validation studies in different populations are necessary to establish these polymorphisms as genetic markers for mastitis.

Acknowledgements

This study was funded by the Science and Technology Development Fund (STDF) in Egypt (grant no. 2047).

References

Abdel-Shafy H., Bortfeldt R. H., Tetens J. and Brockmann G. A. 2014 Single nucleotide polymorphism and haplotype effects associated with somatic cell score in German Holstein cattle. *Genet. Sel. Evol.* **46**, 35–44.

Akhtar S. and Ali S. 1994 Monitoring of bovine health problems of small dairy herds in Islamabad capital territory; design, data and disease frequencies. *Trop. Anim. Health Pro.* **26**, 193–198.

Ali A. K. A. and Shook G. E. 1980 An optimum transformation for somatic-cell concentration in milk. *J. Dairy Sci.* **63**, 487–490.

Ameratunga R., Winkelstein J. A., Brody L., Binns M., Cork L. C., Colombani P. and Valle D. 1998 Molecular analysis of the third component of canine complement (c3) and identification of the mutation responsible for hereditary canine c3 deficiency. *J. Immunol.* **160**, 2824–2830.

Barrio M. B., Rainard P. and Poutrel B. 2003 Milk complement and the opsonophagocytosis and killing of *Staphylococcus aureus* mastitis isolates by bovine neutrophils. *Microb. Pathogenesis* **34**, 1–9.

Botto M., Kirschfink M., Macor P., Pickering M. C., Würzner R. and Tedesco F. 2009 Complement in human disease: lessons from complement deficiencies. *Mol. Immunol.* **46**, 2774–2783.

Bramham J., Thai C.-T., Soares D. C., Uhrın D., Ogata R. T. and Barlow P. N. 2005 Functional insights from the structure of the multifunctional C345c domain of C5 of complement. *J. Biol. Chem.* **280**, 10636–10645.

Cole J. B., Wiggans G. R., Ma L., Sonstegard T. S., Lawlor T. J., Crooker B. A. *et al.* 2011 Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* **12**, 408–424.

Dassault Systèmes BIOVIA 2015 *Discovery Studio Modeling Environment,* Release 4.5. Dassault Systèmes, San Diego.

de Bruijn M. H. L. and Fey G. H. 1985 Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl. Acad. Sci. USA* **82**, 708–712.

Detilleux J. C. 2002 Genetic factors affecting susceptibility of dairy cows to udder pathogens. *Vet. Immunol. Immunop.* **88**, 103–110.

DiScipio R. G. and Hugli T. E. 1989 The molecular architecture of human complement component C6. *J. Biol. Chem.* **264**, 16197–16206.

Falconer D. S. and Mackay T. F. C. 1996 *Introduction to quantitative genetics*, 4th edition. Longman Scientific and Technical, New York, USA.

Firth M. A., Moore D. P., Pei Y., Shewen P. E., Lo R. Y. C., Yoo D. and Hodgins D. C. 2006 Cloning of a gene fragment encoding bovine complement component C3d with expression and characterization of derived fusion proteins. *Vet. Immunol. Immunop.* **114**, 61–71.

Frank M. M. and Fries L. F. 1991 The role of complement in inflammation and phagocytosis. *Immunol. Today* **12**, 322–326.

Fredslund F., Jenner L., Husted L. B., Nyborg J., Andersen G. R. and Sottrup-Jensen L. 2006 The structure of bovine complement component 3 reveals the basis for thioester function. *J. Mol. Biol.* **361**, 115–127.

Goldberg M., Fremeaux-Bacchi V., Koch P., Fishelson Z. and Katz Y. 2011 A novel mutation in the C3 gene and recurrent invasive pneumococcal infection: a clue for vaccine development. *Mol. Immunol.* **48**, 1926–1931.

Hazlewood M. A., Kumararatne D. S., Webster A. D., Goodall M., Bird P. and Daha M. 1992 An association between homozygous C3 deficiency and low levels of antipneumococcal capsular polysaccharide antibodies. *Clin. Exp. Immunol.* **87**, 404–409.

Humphrey W., Dalke A. and Schulten K. 1996 VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38.

Ibeagha-Awemu E. M., Kgwatalala E. M., Ibeagha A. E. and Zhao X. 2008 A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mamm. Genome* **19**, 226–245.

Isaac L. and Isenman D. E. 1992 Structural requirements for thioester bond formation in human complement component C3. Reassessment of the role of thioester bond integrity on the conformation of C3. *J. Biol. Chem.* **267**, 10062–10069.

Jamrozik J. and Schaeffer L. R. 2010 Recursive relationships between milk yield and somatic cell score of Canadian Holsteins from finite mixture random regression models. *J. Dairy Sci.* **93**, 5474–5486.

Janssen B. J. C., Huizinga E. G., Raaijmakers H. C. A., Roos A., Daha M. R., Nilsson-Ekdahl K. *et al.* 2005 Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437**, 505–511.

Kidmose R. T., Laursen N. S., Dobó J., Kjaer T. R., Sirotkina S., Yatime L. *et al.* 2012 Structural basis for activation of the complement system by component C4 cleavage. *Proc. Natl. Acad. Sci. USA* **109**, 15425–15430.

Kimchi-Sarfaty C., Oh J. M., Kim I. W., Sauna Z. E., Calcagno A. M., Ambudkar S. V. and Gottesman M. M. 2007 A silent polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**, 525–528.

Komar A. A. 2007 Silent SNPs: impact on gene function and phenotype. *Pharmacogenomics* **8**, 1075–1080.

Law S. K. and Dodds A. W. 1997 The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6**, 263–274.

MacKerell A. D., Bashford D., Bellott M., Dunbrack R. L., Evanseck J. D., Field M. J. *et al.* 1998 All atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616.

MacKerell A. D., Feig M. and Brooks C. L. 2004 Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415.

Meredith B. K., Kearney F. J., Finlay E. K., Bradley D. G., Fahey A. G., Berry D. P. and Lynn D. J. 2012 Genome-wide associations for milk production and somatic cell score in Holstein–Friesian cattle in Ireland. *BMC Genet.* **13**, 21–31.

Miyagawa H., Yamai M., Sakaguchi D., Kiyohara C., Tsukamoto H., Kimoto Y. *et al.* 2008 Association of polymorphisms in complement component C3 gene with susceptibility to systemic lupus erythematosus. *Rheumatology* **47**, 158–164.

Morgan B. P., Marchbank K. J., Longhi M. P., Harris C. L. and Gallimore A. M. 2005 Complement: central to innate immunity and bridging to adaptive responses. *Immunol. Lett.* **97**, 171–179.

Moroni P., Rossi C. S., Pisoni G., Bronzo V., Castiglioni B. and Boettcher P. J. 2006 Relationships between somatic cell count and intramammary infection in buffaloes. *J. Dairy Sci.* **89**, 998–1003.

Nani J. P., Raschia M. A., Poli M. A., Calvinho L. F. and Amadio A. F. 2015 Genome-wide association study for somatic cell score in Argentinean dairy cattle. *Livest. Sci.* **175**, 1–9.

Ogorevc J., Kunej T., Razpet A. and Dovc P. 2009 Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim. Genet.* **40**, 832–851.

Oviedo-Boyso J., Valdez-Alarcón J. J., Cajero-Juarez M., Ochoa-Zarzosa A., López-Meza J. E., Ravo-Patiño A. and Baizabal-Aguirre V. M. 2007 Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. *J. Infection* **54**, 399–409.

Peakall R. and Smouse P. E. 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**, 2537–2539.

Pearson T. A. and Manolio T. A. 2008 How to interpret a genome-wide association study. *J. Am. Med. Assoc.* **299**, 1335–1344.

Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C. and Ferrin T. E. 2004 UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.

Phillips J. C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E. *et al.* 2005 Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.

Rainard P. 2003 The complement in milk and defense of the bovine mammary gland against infections. *Vet. Res.* **34**, 647–670.

Rainard P. and Poutrel B. 1995 Deposition of complement components on *Streptococcus agalactiae* in bovine milk in the absence of inflammation. *Infect. Immun.* **63**, 3422–3427.

Ricklin D. and Lambris J. D. 2013 Complement in immune and inflammatory disorders: pathophysiological mechanisms. *J. Immunol.* **190**, 3831–3838.

Ricklin D., Hajishengallis G., Yang K. and Lambris J. D. 2010 Complement: a key system for immune syrveillance and homeostasis. *Nat. Immunol.* **11**, 785–797.

Rupp R. and Boichard D. 2003 Genetics of resistance to mastitis in dairy cattle. *Vet. Res.* **34**, 671–688.

Šali A. and Blundell T. L. 1993 Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815.

Sambrook J., Fritsch E. F. and Maniatis T. 1989 *Molecular cloning*: *a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA.

Santos-Valente E., Reisli I., Artaç H., Ott R., Sanal Ö. and Boztug K. 2013 A novel mutation in the complementcomponent 3 gene in a patient with selective IGA deficiency. *J. Clin. Immunol.* **33**, 127–133.

Skattum L., vanDeuren M., vanderPoll T. and Truedsson L. 2011 Complement deficiency states and associated infections. *Mol. Immunol.* **48**, 1643–1655.

Swanson K. M., Stelwagen K., Dobson J., Henderson H. V., Davis S. R., Farr V. C. and Singh K. 2009 Transcriptome profiling of *Streptococcus uberis*-induced mastitis reveals fundamental differences between immune gene expression in the mammary gland and in a primary cell culture model. *J. Dairy Sci.* **92**, 117–129.

Tomlinson S. 1993 Complement defense mechanisms. *Curr. Opin. Immunol.* **5**, 83–89.

Torreira E., Tortajada A., Montes T., de Cordoba R. S. and Lorca O. 2009 3D structure of the C3bB complex provides insights into the activation and regulation of the complement alternative pathway convertase. *Proc. Natl. Acad. Sci. USA* **106**, 882–887.

Walport M. J. 2001 Complement: first of two parts. *N. Engl. J. Med.* **344**, 1058–1066.

Wanger E. and Frank M. M. 2009 Therapeutic potential of complement modulation. *Nat. Rev. Drug Discov.* **9**, 43–56.

Welinder K. G. and Svendsen A. 1986 Amino acid sequence analysis of the glycopeptides from human complement component C3. *FEBS Lett.* **202**, 59–62.

Wijga S., Bastiaansen J. W., Wall E., Strandberg E., de Haas Y., Giblin L. and Bovenhuis H. 2012 Genomic associations with somatic cell score in first-lactation Holstein cows. *J. Dairy Sci.* **95**, 899–908.

Wilmink J. B. M. 1987 Efficiency of selection for different cumulative milk, fat and protein yields in first lactation. *Livest. Prod. Sci.* **17**, 211–224.

Wimmers K., Mekchay S., Schellander K. and Ponsuksili S. 2003 Molecular characterization of the pig C3 gene and its association with complement activity. *Immunogenetics* **54**, 714–724.

Yates J. R. W., Sepp T., Matharu B. K., Khan J. C., Thurlby D. A., Shahid H. *et al.* 2007 Complement C3 variant and the risk of age-related macular degeneration. *N. Engl. J. Med.* **357**, 553–561.

Received 25 February 2016, in final revised form 8 June 2016; accepted 16 June 2016
Unedited version published online: 20 June 2016
Final version published online: 7 February 2017

Corresponding editor: SILVIA GARAGNA