Meta Gene 15 (2018) 1–9

Contents lists available at ScienceDirect

Meta Gene

journal homepage: www.elsevier.com/locate/mgene

Computational analysis of non-synonymous single nucleotide polymorphism in the bovine cattle kappa-casein (CSN3) gene

Jainikkumar B. Patel, Jenabhai B. Chauhan*

P. G. Department of Genetics, Ashok and Rita Patel Institute of Integrated Studies and Research in Biotechnology and Allied Sciences (ARIBAS), Affiliated to Sardar Patel University, New Vallabh Vidyanagar 388121, India

ARTICLE INFO

Keywords:
Non-synonymous SNP
Cattle
CSN3
Computational analysis

ABSTRACT

Background: kappa-casein is candidate gene, present on chromosome 6 of cattle, play an important role in mammary gland and properties of milk.
Methods: In the present study, we used SIFT, SNAP2, PROVEAN and PolyPhen 2 computational tools which are based on different approaches to predict deleterious nsSNPs in coding region of cattle CSN3 gene. Furthermore, protein stability and structure were predicted by I-Mutant 2.0, MuStab, ModPred, ConSurf, PSIPRED, MUSTER, Ramachandran plot, TM-score, FT site and STRING.
Results: Among the 15 nsSNPs, 3 were predicted deleterious by SIFT, SNAP2, PROVEAN and PolyPhen 2. Additionally, R118S, S125A and S176G variant showed decrease in stability, S125A and S176G are highly conserved and role in post translational modification using I-Mutant 2.0, MuStab, ConSurf and ModPred. The 3D structure of CSN3 wild type and mutant were generated by MUSTER and validated via Ramachandran plot, showed 170 residues (90.4%) in favoured region, 14 residues in allowing the region (7.4%) and 4 residues (2.2%) in outer region which indicated good quality model. The TM-score ranging from 0.70 to 0.87 which means two structures was within same fold while higher RMSD values (1.59 to 2.31) showed that the deviation from two structures may change the functional activity of the protein. Results of GROMOS96 indicated that the stability of mutant protein was decreased. Ligand binding site 1 has a SER-125, site 2, ASP-169 and site 3, SER-125. CSN3 protein interacts with 10 different proteins.
Conclusion: The S125A variant of cattle population may serve as a genetic marker for mutational analysis of CSN3.

1. Introduction

Milk components in cattle are quantitative traits and influenced by various environmental and genetic factors. Despite of these, a genetic factor is the most important factor responsible for milk quality and quantity in cattle. Milk is a key source of essential nutrients for lactating calves and important raw materials for human food preparations (Reinhardt et al., 2012). Across the world, people fulfil approximately 13% of their protein requirement from bovine milk and milk products. The cattle milk contains 3–5% protein, out of which 20% is whey protein and 80% is casein (Farrel et al., 2004). kappa-casein (CSN3) constitutes about 25% of the casein fraction of milk and several polymorphisms have been found for this protein. CSN3 casein gene is more polymorphic than alpha S1, alpha S2 and CSN2. These genes occupy region of < 200 kbs on chromosome 6 and it forms a strong gene cluster (Caroli et al., 2009). The CSN3 gene location is 6q31 with 13 kb sequence and divided into 5 exons (Alexander et al., 1988). The highest density of quantitative trait locus (QTL) associated with milk traits was found on BTA6 (bovine chromosome 6) and BTA14 (bovine chromosome 14) (Ogorevc et al., 2009). CSN3 gene is located on BTA6 and play important role in mammary gland.

The casein concentration altered by genetic variants of mammary gland and it has a significant effect on micelle size and properties of milk (Juszczak et al., 2001). CSN3 gene plays the most important role in determining the size of casein micelles and initiate micelle aggregation resulting in curd and cheese production (Threadgill and Womack, 1990). Among various dairy product, cheese production has a great market so it is most important to study CSN3 gene that influences the

Abbreviations: SNP, single nucleotide polymorphism; SIFT, Sorting Intolerant From Tolerant; PolyPhen, phenotype polymorphism; nsSNP, nonsynonymous single nucleotide polymorphism; MUSTER, MUlti-Sources ThreadER; NCBI, National Centre for Biological Information; PROVEAN, Protein Variation Effect Analyzer; SNAP2, Screening of Non-acceptable Polymorphism 2; SVM, Support Vector Machine; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; PDB, Protein Data Bank; PTM, post-translational modification; DAS, Distributed Annotation System; RMSD, root-mean-square deviation of atomic positions

* Corresponding author.
E-mail addresses: pateljainikgt@gmail.com (J.B. Patel), jenabhaichauhan@aribas.edu.in (J.B. Chauhan).

http://dx.doi.org/10.1016/j.mgene.2017.10.002
Received 6 July 2017; Received in revised form 14 September 2017; Accepted 6 October 2017
Available online 06 October 2017

manufacturing properties of cheese (Kishor et al., 2014). Recent development in high-throughput genotyping and next generation sequencing have produced a huge amount of cattle genetic variation data, determining the functional or structural effects of amino acid substitution will be the next greatest challenge in genomics research.

Single nucleotide polymorphisms are common and essential variation in cattle genome and there is a solid correlation between variation and certain economically important traits (Rasal et al., 2015). Missense mutation is also called non-synonymous SNPs (nsSNPs) arises in the coding region, which alters the amino acid configuration which may have an impact on the structure and function of the protein (Wohlrab, 2006). Differentiating deleterious nsSNPs (with significant phenotypic consequences) from tolerant nsSNPs (without phenotypic changes) has great importance in understanding the genetic basis of milk production traits in cattle. The functional and structural relationship knowledge of protein is fundamental to find a molecular basis for genetic traits in cattle. The experimental designing for the mutational changes will be laborious and time consuming. Therefore, it is fundamental and beneficial to carry out the basic work required for mutation design and to know protein properties through computational biology (Nailwal and Chauhan, 2017).

The nsSNPs for the CSN3 gene have not been illustrated to date through computational analysis. Therefore, to explore possible associations with genetic mutation and milk trait variation in cattle, different algorithms like SIFT, PROVEAN, PolyPhen 2 and SNAP2 were used to figure out of highly potential nonsynonymous single nucleotide polymorphisms in coding regions of CSN3 gene that are likely to have an effect on the biological function and structure of the protein. Considering the role played by CSN3 gene in milk production traits in cattle, the study aimed to narrow down the candidate nsSNPs for the CSN3 through computational analysis, which may affect the protein structure and/or function that may serve as an important role in milk production traits.

2. Materials and methods

We used SIFT, SNAP2, PROVEAN and PolyPhen 2 computational tools. Moreover, damaging nsSNPs by these four tools was submitted to I-Mutant 2 and Mutstab for stability analysis. Additionally, conservation analysis (ConSurf), 2D and 3D structure prediction, validation of models, structural effect and protein-protein interaction was done for CSN3 protein. Details of each tool are described below and highlighted in Fig. 1.

2.1. Data set

The SNPs and their related protein sequence of the CSN3 gene were obtained from the NCBI dbSNP database of single nucleotide polymorphism (https://www.ncbi.nlm.nih.gov/snp) and were subjected to various computational analyses for predicting damaging amino acid substitution.

2.2. Non-synonymous SNP functional analysis for CSN3

Four tools, SIFT, SNAP2, PROVEAN and PolyPhen 2 were used to predict the functional context of missense mutations in CSN3 gene.

The freely available online tools were used for sequence and structure based approaches to predict the non-synonymous SNP. The substituted amino acids that alter protein function and phenotypic changes was predicated by SIFT (Sorting Intolerant From Tolerant; http://sift.jcvi.org/) tool. In the current investigation, the identification numbers (rsIDs) of each SNP of CSN3 gene obtained from NCBI were submitted as a query to SIFT for homology searching. Results were obtained as SIFT scores which were classified as damaging (0.00–0.05), potentially damaging (0.051–0.10), borderline (0.101–0.20), or tolerant (0.201–1.00) (Ng and Henikoff, 2003).

SNAP2 (Screening of Non-Acceptable Polymorphism 2) is a tool, developed based on a neural network classification method which is freely available (https://rostlab.org/services/snap2web/). It predicts the effect of nsSNPs on protein function (Hecht et al., 2013). The input query submitted is the protein FASTA sequence and lists of mutants which provided scores of each substitution that can then be translated into binary predictions neutral or non-neutral effect.

PROVEAN (Protein Variation Effect Analyzer; http://provean.jcvi.org/index.php) is used to predict the possible impact of a substituted amino acid and indels on protein structure and biological function. The input query is a protein FASTA sequence along with amino acid substitutions. It analyses the nsSNPs as deleterious or neutral, if the final score was below the threshold score of -2.5 were considered deleterious; scores above this threshold were considered neutral (Choi et al., 2012).

PolyPhen-2 (Polymorphism Phenotyping v2, http://genetics.bwh.harvard.edu/pph2/) is a multiple sequence alignment server, it characterizes the substitution site and calculate the position-specific

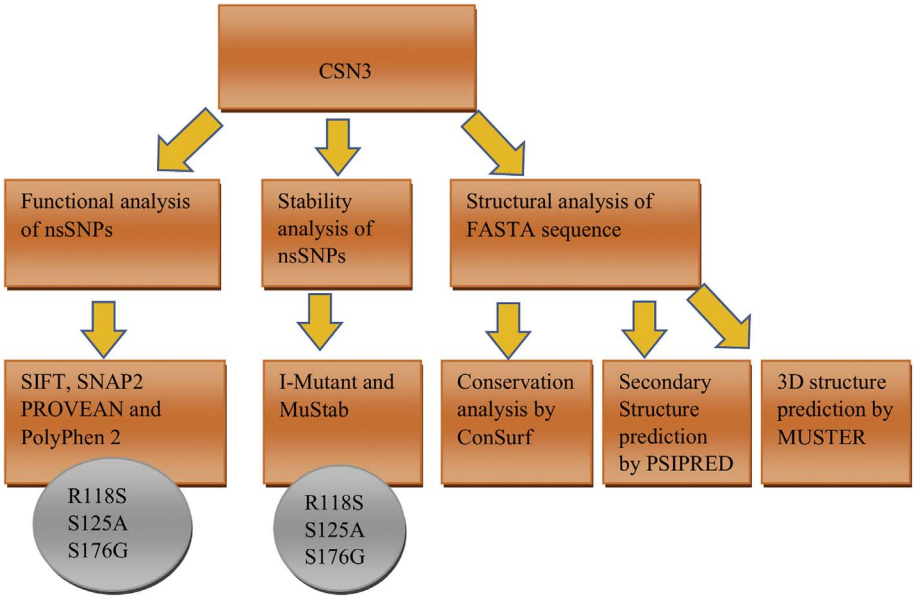

Fig. 1. Diagrammatic representation of computational tools used for analysis of CSN3 gene.