INVITED REVIEW

Role of genome technology in dairy food quality and safety

Sachinandan De

Abstract: Genomics tools, specifically the different affordable Sequencing methods, are rapidly transforming almost every aspect of Biological Sciences including food science. Applications of genomic technologies in food sciences is generally considered a great leap forward with some concerns regarding the need of well-equipped and trained laboratories to discriminate between good and bad micro-organisms in the fermented food stuff. Species identification of raw and finished dairy and animal products has become an important issue regarding the assessment of food composition. Protection against species substitution or admixture in dairy and animal products is of significant importance for reasons relating to consumer preference, public health and government regulations. Milk and milk products are of particular interest, because they are a group of foods that play an important role in feeding the population and are essential for certain groups of consumers (women, children, and the elderly). Milk products are often prone to adulteration practices, since milk is a fairly expensive raw material and can be easily replaced in part by other dairy or non-dairy ingredients. DNA based methods became very popular, since they are considered more reliable as a result of DNA stability under high temperatures, pressures and chemical treatments used during processing of food products. The polymerase chain reaction (PCR) is the most widely used molecular technique, because of its simplicity, sensitivity and reproducibility. Ruminant milk can be easily used as a source of DNA, since it has a large amount of somatic cells, mostly leucocytes but also epithelial cells from the milking mother, which contain genomic DNA suitable for any kind of genomic DNA testing. PCR amplification of various regions of mitochondrial genome, 12S rRNA, growth hormone (GH) gene and PCR RFLP have been reported and proved the sensitivity of DNA based methods and their reliability for species identification in a wide range of raw and finished animal food products. Complete genome sequencing of bacterial isolates (pathogenic or lactic acid bacteria) has started to replace contemporary popular molecular methods, like ribotyping, MLST and Pulse Field Gel Electrophoresis (PFGE), as subtyping method of choice. Whole Genome Sequencing can able to route of transmission in case of increasingly smaller outbreaks.

Keywords: Consumers, Food safety, Genomics, Milk and milk products

Introduction

The production of cereal grain, milk, meat and other animal products throughout the world has been greatly influenced by the genetic improvement of plants and animals through breeding. Because of high intensity of manmade selection process the genetic variability of the domestic animals are reduced over the period of time and we are getting maximum output from the desired animals and plant. Genomics is playing a very crucial role in estimating the production capabilities of a particular animal or plant species. Along with production capability, complete genome information also required for evaluation of the global adaptability and biodiversity of an animal species. Advanced genomic tools help in identifying the segments of the genome responsible for higher production or adaptation. It can also improve our understanding of microevolution through a better understanding of natural selection, mutation, and recombination. The knowledge gained from genomics when it is properly applied, will provide breeders, farmers and indeed mankind with the genetic variation needed to fuel further improvement. The variation available through traditional means is unlikely to produce the higher
yielding varieties with improved resistance to disease and environmental stresses. Improvements in milk quality and nutritional value will also be facilitated through modifications of the gene content guided by acquired knowledge about gene function.

**Complete genome sequence of food animals**

Scientific Research organizations became interested in the genomes of the animals to increase the ability of animal to provide high quality, low cost and safe animal products to the consumer. Mouse and human genome sequence are used as a reference genome sequence for all mammalian species. The complete reference genome information and competent data interpretation revolutionized the food science and other area of researches. Different affordable sequencing methods, are rapidly transforming in almost every aspect of Biological Sciences including food science. With the advance in genome technologies, the genome of the several livestock species (chicken, cow, sheep, pig, horse and rabbit) has been sequenced, and efforts are being made to update and improve the information annotated to them. In present day complete genomes can be sequenced by different methods, in practice all of them result in a pool of millions of short (75–150 bp) or long (>500 bp) sequence reads. Human genome sequence was delivered in 2001, after few years the first sequence of the cow (2004), chicken (2005), horse (2007), pig (2010), rabbit (2014) and sheep (2014) genomes were available. As the costs of WGS have become more affordable, it is now feasible to describe genetic variability in a population by sequencing key genetic contributors. As a result of the genome sequencing projects, we have been able to measure the total size of the genome, which is specific to each species. The human genome is slightly longer among mammalian species (3.1 Gbp). The longest plant genome so far sequenced is the loblolly pine tree (*Pinus taeda*) which spans 23.2 Gbp (Blasco and Pena 2018). Comparative analysis of genomic sequence in mammalian species allows the identification of conserved regions of the genome from diverse species. Such comparative analysis has identified conservation of gene content and gene order between the genomes of almost all ruminant animals, including cattle, buffalo, sheep and goat.

**Milk genes**

Milk is an important component of a balanced diet and contains a number of valuable nutritional constituents. Bovine milk is a balanced liquid food (87% water) which contains an average of 13% total solids and about 9% non-solids-fat. Milk provides high-quality proteins with wide range of nutritional, functional, and physiological activities and bioactive peptides. The chief milk protein is casein which constitutes 80% (w/w) of total milk protein, the insoluble fractions and rest 20% (w/w) are whey proteins, the soluble fractions. Milk also contain some minor proteins: serum albumin, immunoglobulins, lactoferrins, transferrins, calcium binding proteins, prolactins and folate binding proteins. The casein protein is encoded by four tightly linked genes spanning 250 Kb of genomic DNA segment mapped on sixth chromosomes in goat and cattle, and in seventh chromosome in buffalo in the orders of: *CSN1S2* (αS2), *CSN2* (β), *CSN1S1* (αS1) and *CSN3* (κ). Alpha S1, alpha S2 and beta casein are calcium sensitive. Alpha S1 (CSN1S1) casein is of around 17.5 Kb with 19 exons. Alpha S2 (CSN1S2) is of 18.5 Kb with 18 exons and 17 introns. Beta casein (CSN2) on the other hand is around 8.5 Kb and has nine exons of which seventh exon being the longest. Calcium insensitive kappa casein (κ-CN) is a 13 kb gene with five exons and four introns and is thought to be genetically divergent from other calcium sensitive caseins (Caroli et al., 2009). Whey protein is a collection of globular proteins with a high level of α-helix structure and the acidic-basic and hydrophobic-hydrophilic amino acids are distributed in bovine chromosome 5 and 11. Alpha-Lactalbumin (α-LA) and beta-lactoglobulin (β-LG) are the predominant whey proteins and comprise about 70–80% of the total whey proteins (Davoodi et al., 2016).

**A1 and A2 beta casein based milk controversies**

Milk is considered as complete food and it is a part of diet since long time. Milk protein constitutes about 80% of total protein which is mainly casein (alpha S1, alpha S2, beta and kappa). Beta-casein, representing about 35% of total caseins, which exists in two different forms (A1 and A2). This is how the milk is being distinguished as A1 and A2 milk. A1 and A2 milk protein contains 209 amino acids, the only difference is at the 67th amino acid where A1 has histidine and A2 has proline. In 1992, when scientists in New Zealand established a correlation between the prevalence of type-1 diabetes and the type of milk consumed, it led to the discovery of the so-called A1 and A2 types of milk. Milk containing A1 beta-casein, at the time of digestion in the small intestine, releases a bioactive peptide called beta-casomorphin-7 (BCM-7). This is an opioid and suspected to have an inhibitory effect on immune function and also suspected to induce type-1 diabetes, heart disease, infant death and autism. On the other hand A2 milk do not release BCM-7. Around 65% of HF cows in North America produce A1 milk at the same time, 90 per cent cows of the same breed in Germany produce A2 milk. In India, 98% of our cows and all buffaloes produce A2 milk. In 2000, a New Zealand’s A2 Corporation Limited started selling A2 milk after genetic testing was done, around 2003, Food Standards Australia New Zealand, a bi-national government agency, made sure to print the health warnings on A1 milk packets. Keith Woodford (2006) in his book entitled ‘Devil in the Milk’ also linked A1 beta-casein intake to Type 1 diabetes. European Food Safety Authority in 2009, on the other hand declared that no cause and effect relationship could be established between the dietary intake of BCM-7 and various diseases. Such controversies are still there regarding A1 and A2 milk.
A PCR based method was developed in Animal Genomics Laboratory of Animal Biotechnology centre to detect the A1 and A2 beta casein variant forms in cattle and buffalo milk. In Indian cattle milk A1 beta casein is represented in heterozygous A1A2 type as well as in A1A1 type. Indian Buffalo is exclusively A2 type. We have validated this test in more than 1000 samples (blood and milk). At DNA level, a mutation in the DNA sequence coding for the beta casein protein at nucleotide position 200 has resulted in the replacement of a cytidine base with an adenine base. Thus, the triplet codon affected by this change codes for histidine (CAT) rather than for proline (CCT) corresponding to amino acid position 67 of the protein. The PCR test is based on the detection of A or C nucleotide in the beta casin coding region (Marangoni, et al, 2018).

Milk authentication

Authenticity of food is an issue that is growing in awareness and concern worldwide. Food adulteration not only influences the food quality but also poses harmful health effects. Since the modern food supply and demand has increased the risk of food fraud has also increased. Food fraudulent refers to deliberate substitution, addition, tampering, or misrepresentation of food, food ingredients, or food packaging; it also includes false or misleading statements made about a product for economic gain. Thus testing and detection of adulterant various food products are required for value assessment and to assure consumer protection against fraudulent activities. The methods for identifying the species origin in milk and milk products are mainly based on polymerase chain reaction (PCR), DNA and protein/peptide due to their high specificity and sensitivity and rapid processing time. Even the presence of Indian or exotic breeds can also be identified with DNA based analysis. The variability of DNA at the species and target tissue limits DNA-based methods for the exact quantification of percentages of different species in milk and milk products. Immunological methods like enzyme linked immunosorbent assay (ELISA) make use of the antigen-antibody interactions, has been the most widely used technique in detecting food authenticity because of its specificity, simplicity and sensitivity (Ballin et al., 2009).

Milk adulteration is an important issue globally. The addition of cow’s milk to sheep’s, goat’s or buffalo’s milk or other dairy products and labeling them as pure cow or pure buffalo milk is seen ever where. To assure consumers of accurate labeling, it is necessary to prove the authenticity of labels, using fast, reliable and sensitive methods for species identification. PCR based analytical methods are the most reliable and sensitive technique. PCR-based methods to analyze milk authenticity are particularly focused on the analysis of mitochondrial DNA sequences, because its sequence is highly conservative within different animal species and gives high copy number in animal cells, increasing the specificity of the method (Zarei et al 2016). The selective amplification of the conserved region of mitochondrial cytochrome b gene specific to mammals was used for species identification (Bos, Ovis, Capra, Bubalus etc) in dairy products by PCR or PCR-RFLP. The amplification of conserved regions of mitochondrial 12S and 16S rRNA genes among different animal species was also tested for the detection of milk adulteration with cow milk. Bottero et al. (2003) developed a triplex PCR for identification of cow, goat and sheep milk in cheese samples, the sensitivity of the method was 0.5% of bovine in caprine milk. Tortorici et al (2016) developed a duplex PCR for the amplification of cow, ewe and goat DNA sequences specific to 12SrRNA gene. Substitution of buffalo with bovine milk in Mozzarella cheese was tested by species identification of bovine specific mitochondrial cytochrome b gene and mitochondrial cytochrome oxidase subunit I gene. Real-time PCR, a method with high throughput and reduced analysis time has been in use for such studies. The identification can be based on species-specific Taqman or fluorescent probes. Qadruplex real-time PCR using Taqman probes was developed by Rentsch et al. (2013) for simultaneous

### Table 1 Animal genome sequences available today

<table>
<thead>
<tr>
<th>Animal/species</th>
<th>Size, Gbp</th>
<th>Platform</th>
<th>Sequencing center/publication</th>
<th>Submission date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffalo (Bubalus bubalis)</td>
<td>2.66</td>
<td>PacBio Sequal</td>
<td>University of Adelaide, Roseworthy, Australia</td>
<td>16-01-2019</td>
</tr>
<tr>
<td>Cattle (Bos taurus)</td>
<td>2.72</td>
<td>PacBio</td>
<td>Cattle Genome Sequencing</td>
<td>19-11-2015</td>
</tr>
<tr>
<td>Hereford (Bos indicus)</td>
<td>2.67</td>
<td>SOLID</td>
<td>International Consortium</td>
<td>2012</td>
</tr>
<tr>
<td>Nelore breed (Ovis aries)</td>
<td>2.61</td>
<td>PacBio</td>
<td>Genoa Biotecnologia SA(10.1093/jhered/esr153)</td>
<td>20-11-2015</td>
</tr>
<tr>
<td>Sheep (Capra hircus)</td>
<td>2.92</td>
<td>PacBio</td>
<td>USDAARS</td>
<td>24-08-2016</td>
</tr>
<tr>
<td>Pig (Sus scrofa)</td>
<td>2.50</td>
<td>PacBio</td>
<td>Swine Genome Sequencing Consortium (SGSC)(10.1111/age.12548)</td>
<td>02-07-2017</td>
</tr>
<tr>
<td>Chicken (Gallus gallus)</td>
<td>1.14</td>
<td>PacBio RSII</td>
<td>Genome Reference Consortium GRCg6 scaffold</td>
<td>02-02-2018</td>
</tr>
</tbody>
</table>

(Adopted from: Gatew and Tarekeng 2018)
determination of cow, sheep, goat and buffalo milk in milk, cheese and yogurt samples. A quadruplex (cow, goat, sheep and buffalo) real-time PCR based on DNA intercalating fluorescent dye and HRM analysis has been also developed by Agrimonti et al. (2019) with detection limit of 0.1% for cow species in home-made cheese mixtures.

Genomics in food microbiology

Genomics in food microbiology has provided the understanding of biological similarities and dissimilarities as well as evolutionary relationships of food microbes. It helps in the assessment of microbial diversity. Genetic tools like PCR, DNA hybridization and GeneChip have been used in genomics. The bacteria of interest to food microbiology can be divided into infectious agents, causes of food borne intoxication, spoilage and processing aids. Food biotechnology is mainly concerned with application of live microorganisms like yeasts, fungi, bacilli and lactic acid bacteria in industrial processes (Gill 2017). There is a huge influence of food microbes on human health, for example, in pre- and probiotics research. The emerging techniques, such as single nucleotide polymorphism (SNP)-based techniques that use whole-genome sequencing (WGS) offer a resolution that was previously not possible. Before the introduction of next generation sequencing techniques in 2005, DNA sequencing has been carried out using Maxam and Gilbert (1977) and Sanger et al. (1977) sequencing methods (Heather and Chain, 2016). Due to low speed, expensive and time consuming problems of these methods, a new high throughput method known as NGS was introduced. With WGS, data can be used to elucidate phylogenetic relationships; disease-causing line ages can be tracked and monitored over time, virulence gene detection, antibiotic resistance gene profiling, synten comparisons, mobile genetic element identification, and geographic attribution (Ronholm et al., 2016). Second- and third-generation sequencing platforms have advanced microbial whole-genome sequencing (WGS) to the point where it has become available for routine use in research and reference laboratories and etiological/genetic element identification, and geographic attribution (Ronholm et al., 2016). Second- and third-generation sequencing platforms have advanced microbial whole-genome sequencing (WGS) to the point where it has become available for routine use in research and reference laboratories and etiological/epidemiological investigations. More recent developments in sequencing, such as those that are capable of sequencing vast numbers of microbial genomes. As of February 21, 2018, the GenomeTrakr Network has submitted sequence data for 16,864 L. monocytogenes; 105,330 S. enterica; and 38,347 E. coli and Shigella. WGS analysis is a powerful tool for enforcement activities that provides specific and actionable information that empowers stakeholders to supply safe and sanitary food products (Pightling et al., 2018, Allard et al., 2018).

Conclusions

Using whole genome sequencing data collaborations between clinical and regulatory agencies have allowed for outbreaks to be identified through retrospective analysis of isolates collected during regulatory actions. Further improvements in genome sequencing technology and bioinformatics tools will allow food safety researchers to concentrate on asking critical biological questions rather than grappling with complex data sets.

References