Chapter 1

Introduction
The Indian subcontinent is currently populated by more than one billion individuals with huge cultural, linguistic, ethnic diversity. Along with this diversity; individuals within this population have diverse food habits; vary in their digestive capabilities and susceptibility of various diseases. Wheat is the main cereal of the Northern region of India. In North India, people prefer flat bread of wheat flour. However, in the South and Northeast of the Indian people prefer rice as a main food. On an average, North Indian people have an intake of 25 to 30g of gluten per day, while rest of the Indian population have 10 to 20g gluten intake per day (Price 2005, Rajpoot and Makharia 2013). Gluten is one of the major constituent of the total dietary proteins present in various cereals like wheat, barley, oat, sorghum, maize, barley and rye (Gokhale et al., 2003). As time advances, various studies had enriched our understanding about the digestion, absorption of various food components. However, digestion of various types of dietary proteins and role of the host digestive system is still debatable (Tasse et al., 2010). Gluten comprises of water insoluble prolamine and water soluble gliadin, which together comprises 80% of cereal protein contents (Wiser 2007). Gluten protein family comprises of, hundred different types of protein components that are available either in the form of monomers (single-chain polypeptides) and polymers (multiple polypeptide chains linked by disulphide bonds) proteins (Wrigley and Bietz 1988). Glutens are unique as they contain a high percentage of proline and glutamine and a very low percentage of charged amino acids. Majority of the population (>70%) utilizes them and likewise other proteins, they were supposed to be digested by host proteolytic enzymes into oligopeptides, single amino acids. However, many studies have reported that digestion of gluten by host enzymes are quite difficult (Helmerhorst et al., 2010). High content of proline and glutamine are making gluten resistant to cleavage by the major human gastrointestinal digestive enzymes (Wiser 2007). Various studies have shown that the pepsin or trypsins are not capable to cleave the peptide bonds C-terminal to these residues (glutamine and proline) and partial digestion of gluten generates multiple peptides (Helmerhorst et al., 2010). Partially digested peptides reach the duodenum, e.g. a 33-mer peptide from Ūgliadin and a 26-mer peptide from Ñ-gliadin, are indicators for the inability of the human body to systematically digest gluten (Helmerhorst et al., 2010). Allergies could either be induced either by pentapeptide (Gln-Gln-Gln-Pro-Pro) or by terapeptide (Gln-Gln-Gln-
Pro and Pro-Ser-Gln-Gln) of these peptides that could result into celiac diseases. These peptides were modified by the transglutaminase enzyme present in duodenal mucosa and a negative charge was introduced into the peptide, which enhances class II MHC binding on antigen-presenting cells. That leads to T cell activation, which results in the destructive immunological responses in the proximal intestine of celiac disease (CD) patients (Jabri and Sollid 2006, Koning et al., 2005, Korneychuk et al., 2015). Moreover, CD is a T-cell related disease in which gliadin peptides activate immune cells of the lamina propria of small intestine and react with T lymphocytes to initiate adaptive Th1 response. Th1 response enhances the population of interferon gamma (IFN-γ) and interleukin-15 (IL-15), leading to intraepithelial lymphocytes toxicity and finally onset of CD (Serena et al., 2015). Gastrointestinal symptoms are seen in all CD patients, most of them are uncharacterized and carrying some other diseases like osteoporosis, short stature, anemia, convulsions, idiopathic ataxia, depression, skin manifestations, neurologic diseases, infertility, osteopenia, isolated liver diseases, diabetes, epilepsy, cerebral calcification and dental enamel defects. Common physiological symptoms of CD in adult patients are diarrhea, weakness, weight loss, vomiting, loose stools, a distended abdomen. However, most surprisingly celiac disease is developed only in a small portion of the total population (1-5%), who are utilizing these biomolecules on a daily basis. It raises a major concern about the host resistance mechanisms toward the onset of CD in majority of the gluten utilizing population. It provides a clue about the existence of another mysterious mechanism that is protecting the majority of gluten eating population form occurrence of celiac diseases. Many studies have shown an association of onset of celiac diseases with genetic polymorphisms in HLA-DQ2/DQ8 genes of humans (Tjon et al., 2010). Though it has been confirmed that proteolytic degradation of protease-resistant domains in gluten appears to require enzymatic cleavage specificities (specially, endoproteases) (Rizzello et al., 2007). These enzymes should have a property of proteolytic cleavage at proline present at N-terminal or C-terminal or within their amino acid sequence. These enzymes were not willingly accessible in the repertoire of mammalian digestive enzymes, so how this polymorphism could responsible alone for this gluten sensitivity. These endoproteases was identified only of the microbial origin (Helmerhorst et al., 2010). Prolyl endopeptidase (PEP) enzymes are also reported in
different microbes like *Deinococcus radiodurans*, *Shewanella oneidensis*, *Trichodesmium erythraeum*, Nostoc sp., *Nostoc punctiforme*, *Flavobacterium meningosepticum*, *Myxococcus xanthus*, *Thermobifida fusca*, *Sphingomonas capsulata* and *Rothia aeria* (Venalainen et al., 2004b, Zamkachari et al., 2011). In the last few years, various prolyl endopeptidases (PEPs) enzyme has been discovered and identified for their active potential therapeutic agent. Prolyl endopeptidases could easily cleave the gluten derived peptides and have an ability to digest gluten in the gastrointestinal tract. PEPs were used as a drug in the oral treatment of celiac disease. Therefore, how the role of microbes in the progression of this disease could be neglected. Even a strong correlation has been observed between gluten sensitivity and human gut microbial community dynamics (Sanz 2010). Similarly, oral cavity microorganisms were producing endoproteases that could specificity cleave the peptide after a glutamine residue (Rizzello et al., 2007). All these results strongly indicated the possible role of the gut microbiome in digestion of gluten. It was invented that human survives with the constant association of different microbes. These microbes harbor on different body sites, including surfaces, cavities and even the cells of the human body. A huge number of different microbial communities are at least tenfold more in comparison to cells of the human body and the number of predicted unique encoded genes to be at least hundred folds greater than the number of genes present in the human genome (Fierer et al., 2012). Human microbiota/microbiome colonization begins with birth and shows maximum diversity in the healthy adult age e.g. new born baby vaginal microbes are more matched with the mother’s vagina microbiota. In the steady ecosystem, healthy adult life carries more complex microbiota and is dominated by *Bacteroidetes* and *Firmicutes* phyla (Rajili and Stojanovi et al., 2009). Microbes have been identified from various anatomical body sites like intestine, oral cavity, gut, dental plaque, ear, skin, armpits and vagina (Cho and Blaser 2012). The important taxa of human microbial communities are *Actinobacteria*, *Proteobacteria*, *Bacteroides*, *Firmicutes*, *Fusobacteria* and *Verrucomicrobia*. *Firmicutes* contain low G+C and it is a Gram-positive bacteria. Many microbiologists have discovered a lot of specific bacterial taxa related to different human diseases.

Few years back, Human Microbiome Project (HMP) (http://commonfund.nih.gov/hmp/), MetallIT (http://www.metahit.eu/), and other human microbes related major research
projects provide a new area of research for diverse groups of researcher and medical professionals. They easily understand and believe in the signatures and structures of different microbial communities that association and involvement in human health. Gut microbes are involved in many metabolic pathways by which microbes provide essential metabolites for the human bodies. Some of the gut microbes associated metabolites are SCFAs, choline metabolites, bile acids, phenolic, indole derivatives and benzoyl, and phenyl derivatives. Microbes are also involved in the regulation of environmental condition within the body habitats and influence tissue development (Gill et al., 2006, Pflughoeft and Versalovic 2011, Spor et al., 2011). Gut bacteria also protect the host intestine by the colonization resistance (Stecher and Hardt 2008). Pathogenic bacteria colonization and their growth also inhibited by the gut bacteria toxic metabolites and volatile fatty acids (Beaud et al., 2005). Human small intestinal epithelial tight junction protein is regulated by L. plantarum and provides protection (Ulluwishewa et al., 2011, Zhang et al., 2015). Colonization of human gut microbes also induce a prominent response of the immune system of the distal colon to the secretion of IgA, that play a fundamental role in the small intestine gut bacterial community's regulation (Tsuji et al., 2008). Permanent gut microbes are associated with reductive reactions like, acetogenesis, methanogenesis, sulfate reduction, and nitrate reduction.

Human gut microbiome is a dense and playing a vital role in metabolizing dietary constituents (Sonnenburg et al., 2005). Other than this, human gut bacteria were also associated with different diseases due to abnormal changes occurs in the gut ecosystem (Zhang et al., 2015). Like symbiosis, some bacteria are also showing a dysbiosis and cause many diseases into the host like Clostridium difficile is a pathogenic bacterium and it is a permanent member of the human microbiome and it is present in huge numbers (McFarland 2008). Likewise, some other wide range of diseases which are caused due to different bacteria, e.g., antibiotic associated diarrhea, bacterial vaginosis, crohn disease, gingivitis, irritable bowel syndrome, obesity and psoriasis (Fierer et al., 2012, Frank et al., 2011, Pflughoeft and Versalovic 2011). Human gut microbiome is mainly composed of bacterial phyla, as well as several archaeal and eukaryotic species. With up to $10^{12}$ cells per gram of feces, the bacterial abundance is estimated to reach 1000 operational taxonomic units (OTUs) per individual, 70% to 80% of the most dominant ones being
subject-specific (Tasse et al., 2010). However, only 20% of the bacterial species have been successfully cultured so far (Eckburg et al., 2005) and the majority of the microbes remains uncultured. This huge percentage of uncultured microbes, posses a vast number of untapped gene pool which can be explored for novel genes/pathways for various biotechnological applications. This untapped gene pool can be explored either by developing new strategies for using a culture independent tool. Among them, culture independent metagenomic analysis has helped to explore their untapped gene pool to provide gene catalogs for various protein families involved in the predominant functions of the human gut microbiome like carbohydrate catabolic enzymes, stress resistant genes and antibiotic resistance genes (Tasse et al., 2010). Metagenomic approaches whole-genome shotgun sequencing and assembly technique is used for the study of dense microbial communities. To date, 70 divisions of bacteria and 13 divisions of Archaea were known (Karlsson et al., 2013). A phylogenetic evaluation of the microbe’s census has given information for interpreting the functional predictions from metagenomic data. Taking all these facts into consideration, current study was designed to decipher the human gut microbiome for glutenase coding genes. In the present studies, human gut microbiome was explored with culture independent metagenomic approach. The human gut microbiome was explored for its microbial community structure with 16S rRNA gene sequencing and CD specific microbial markers were discovered. A number of glutenase coding genes were deciphered using functional and sequence based metagenomic approach.