Chapter 2
Review of Literature

2.1 Human intestine and associated microbes:

The mucosal surface of the human gastrointestinal tract (GIT) is about 200 –300m^2 and is colonized heavily by around by $10^{13–14}$ bacteria representing several species. According to Savage (Savage, 1977), gastrointestinal microflora can be categorized into autochthonous flora (indigenous flora) and allochthonous flora (transient flora). Autochthonous microorganisms inhabit specific habitats, i.e., physical spaces in the gastrointestinal tract (GIT), whereas allochthonous microorganisms colonize only in abnormal and disturbed states of the gut. Majority of the pathogens are allochthonous microorganisms however, some pathogenic microbes have been found to be autochthonous and live in harmony and exhibit infectivity only when intestinal equilibrium is altered (Trenschel *et al*., 2000). Various factors like pH of the intestine, redox potential, bacterial adhesion and cooperation, mucin secretion, availability of nutrients, peristalsis, etc., decide the composition of microbes in the gut. The population of microbes seems to be low in the upper part of the intestine due to low pH and reduced peristaltic activities. Acid-tolerant lactobacilli and streptococci mainly predominate in the upper tract of the intestine. Distal small intestine (ileum) is characterized by the presence of diverse microbes due to lower oxidation-reduction potentials and decreased peristalsis at the site (Tannock, 1983). The colon harbors numerous microbes that mainly comprise of obligate anaerobes.

Bacterial phyla Firmicutes and Bacteroidetes mainly constitute microbial composition of healthy gut microbiota. Bacterial members representing phyla Actinobacteria, Proteobacteria, Verrucomicrobia and Fusobacteria also reside in human gut, but in lower proportions (Eckburg *et al*., 2005).

Streptococcus has been found to be prevalent in the distal oesophagus, duodenum and jejunum (Justesen *et al*., 1984; Pei *et al*., 2004). The genus *Helicobacter* dominantly resides in the stomach and crucially influences the entire microbial composition of the gastric flora. It has been reported that when *Helicobacter pylori* resides as a commensal there is rich diversity of various microbial genera like *Streptococcus, Prevotella, Veillonella* and *Rothia* (Andersson *et al.*, 2008; Blaser, 1999). However, on acquiring pathogenic phenotype, *H. pylori* significantly administers the reduction of other microbial members. Large intestine harbors several bacterial members that are crucially associated with homeostasis. Firmicutes and Bacteroidetes are most prevalent in the large intestine. It has been suggested that the Firmicutes: Bacteroidetes ratio crucially governs the manifestation of disease states (Ley *et al*., 2006). Pathogenic bacterial species like *Campylobacter jejuni, Salmonella enterica, Escherichia coli* and *Bacteroides fragilis* have also been reported to be occurring in low traces in human colon (Gillespie *et al*., 2011). Abundance of bacterial genera like *Bacteroides, Prevotella* and *Ruminococcus* and reduced proportions of bacterial members representing bacterial phylum Proteobacteria signify healthy gut microbiota (Hollister *et al*., 2014). Members of genera *Bacteroides, Bifidobacterium, Streptococcus, Enterobacter, Enterococcus, Clostridium, Lactobacillus* and *Ruminococcus* are found in high proportions in the lumen whereas, *Lactobacillus, Enterococcus and Akkermansia* are frequently encountered in mucus layer and intestinal epithelial lining (Swidsinski *et al*., 2005).

2.2 Factors influencing the gut microbial composition:

Microbial composition of human intestine remains stable during adulthood (Ley *et al.*, 2006; Vanhoutte *et al.*, 2004; Zoetendal *et al.*, 1998) however, subtle differences always exist between every individual (Eckburg *et al*., 2005; Ley *et al*., 2006). Several factors vitally govern gut microbial composition and have been discussed below:

2.2.1 Host genetics:

Host genetics is an important determinant of gut microbial composition and behavior. Zoetendal *et al*. (2001) inferred that monozygotic conditions signifying genetic likeness

result in similar patterning of gut microbiota. Furthermore, it has been suggested that monozygotic pairs of individuals display more similarities in gut microbial composition than dizygotic twins (Van de Merwe *et al*., 1983; Stewart *et al*., 2005). Genetic factors have also been confirmed to shape gut microbiome significantly (Ley *et al*., 2005).

2.2.2 Birth delivery mode:

The mode of delivery is an imperative factor that decides microbial framework of human intestine. Gut composition of cesarean babies have been found to display slower diversification and lack anaerobic species like *Clostridium* (Gronlund *et al*., 1999).

2.2.3 Geographical impacts:

Geographic locations strongly correlate with discrepancies in gut microbial composition between individuals from different locations. It has been observed that gut microbiome considerably varies between individuals from westernized developed countries and members from developing countries. *H. pylori* population has been found to be more predominant in members from developing countries than developed countries (Genta *et al*., 1995).

2.2.4 Influence of ageing:

Gut microbiota varies between young and elderly individual adults (Woodmansey, 2007). Age is a crucial factor that influences gut microbial constitution. Altered physiological state, significant changes in dietary habits and reduced motility of the intestine in elderly aged individuals contribute for the observed discordances (Woodmansey, 2007; Mueller *et al*., 2006). Elderly individuals display low proportions of *Bifidobacterium*, *Bacteroides* and other amylolytic bacteria (Woodmansey, 2007; Mueller *et al*., 2006). Certain facultative anaerobes like *Fusobacterium*, *Clostridium* and *Eubacterium* have been found to occur frequently in aged individuals (Woodmansey, 2007; Mueller *et al*., 2006).

2.2.5 Influence of diet:

Diet is probably the most vital factor that shapes gut microbial constitution. Dietary habits regulate and control human intestinal health both in beneficial and detrimental manners. It has been interesting to note that microbial composition of formula fed infants vary significantly. Breast-fed infants exhibit higher shares of lactic acid bacteria and bifidobacteria

(Balmer and Wharton, 1991; Harmsen *et al*., 2000). It has also been observed that obese people, who adopt fat-restricted low calorie diet, often display higher frequencies of bacterial members representing Bacteroidetes phylum and a low proportion of members from the Firmicutes phylum (Ley *et al*., 2006).

2.2.6 Impact of antibiotics:

Prolonged usage of antibiotics has been reported to alter gut microbial constitution (Jernberg *et al*., 2007; Lofmark *et al.*, 2006) in a significant manner. Antibiotic treatment causes rapid alteration in gut microbiome. Recently, it has been observed that *Bacteroides* community of human gut never regains its original composition after being exposed to 7-day clindamycin treatment (Jernberg *et al*., 2007). Functionalities and performance of gut microflora also change considerably after antibiotic treatment (Norin, 1997).

2.2.7 Prebiotics and probiotics:

Probiotics are defined as live microorganisms which when administered in adequate amount, into the host system, bestow a health benefit on the host (Fuller, 1986). Lactic acid bacteria (LAB) and bifidobacteria mainly serve as probiotics rendering beneficial effects on human health. Prebiotics like inulin and fructooligosaccarides enhance the growth and viability of bifidobacteria and lactobacilli in the gut (Van Loo, 2004; Gibson and Roberfroid, 1995). Prebiotics and probiotics have been reported to be associated with several beneficial functions that include protection against allergy development (Abrahamsson *et al*., 2007), inflammatory bowel diseases (IBD) (Sartor, 2004), irritable bowel syndrome (IBS) (Kajander *et al.*, 2008; Sartor, 2004) and acute diarrhea (Canani *et al*., 2007).

2.3 Functional aspects of human gut microbiota:

Gut microflora maintains harmony with human intestine and confers several protective and ameliorating effects on intestinal health. Microbial members of human gut derive essential nutrients from its host and in lieu facilitate digestion and help in maintenance of intestinal balance (Sonnenburg *et al*., 2005). Several useful effects of gut microbial members have been discussed below:

2.3.1 Nutrient metabolism:

Dietary carbohydrate components of human diet serve as nutrient sources for the members of gut microbiota. Gut

-associated microbes representing several bacterial genera like *Bacteroides*, *Roseburia*, *Bifidobacterium*, *Fecalibacterium*, and *Enterobacteria* produce short chain fatty acids (SCFA) such as butyrate, propionate and acetate that act as energy sources for human host (Macfarlane and Macfarlane, 2003; Sartor, 2008). Butyrate has also been reported to prevent the accumulation of toxic products like D-lactate (Bourriaud *et al*., 2005) in human gut. Various species of *Bacteroides* encode Carbohydrate-Active enZymes (CAZymes) that facilitate proper breakdown of non-degraded carbohydrate components of human diet (Cantarel *et al*., 2012). *Oxalobacter formigenes*, residing in human gut, degrades oxalate which is produced as a product of carbohydrate fermentation. Certain members of the genera *Lactobacillus* and *Bifidobacterium* also degrade oxalate efficiently. Oxalate increases the risk of formation of stones in the kidney. Thus, the concerned gut microbes reduce the hazard of stone formation in kidney (Magwira *et al.*, 2012; Sidhu *et al.*, 1998).

Human gut-associated microbes have also been reported to synthesize vitamin K and crucial components of vitamin B (LeBlanc *et al.*, 2013). Various members of the genus *Bacteroides* are known to produce conjugated linoleic acid (CLA) - a compound with well-established antidiabetic, antiobesogenic and immunomodulatory activities (Baddini Feitoza *et al.*, 2009; Devillard *et al.*, 2007; Devillard *et al.*, 2009).

Polyphenols (phenolic compounds) are efficiently degraded by bacterial members of human gut (Marin *et al.*, 2015). Gut microbiome has also been reported to effectively raise the concentrations of citric acid, pyruvic acid, malic acid and fumaric acid in serum which eventually facilitates metabolism of human host (Velagapudi *et al*., 2010).

Bacteroides thetaiotaomicron, an important microbial constituent of human gut, has been reported to induce the production of colipase - a protein co-enzyme that facilitates proper hydrolysis and digestion of lipids by the enzyme pancreatic lipase (Hooper *et al*., 2001). Bacterial members like *Bacteroides intestinalis*, *Bacteroides fragilis* and *Escherichia coli* have been reported to be efficient convertors of primary bile acids into secondary bile acids like deoxycholic and lithocolic

acids in human colon (Fukiya *et al*., 2009).

2.3.2 Xenobiotic and drug metabolism:

Members of human gut microflora efficiently degrade and metabolize xenobiotics and drugs and thus, aid in restoring the normal balance of the gut (Clayton *et al*., 2009). It has been observed that β-glucoronidase, a microbial metabolite, disrupts anticancer drug irinotecan that has adverse effects like diarrhea and inflammation (Wallace *et al*., 2010). Furthermore, up-regulation of cytochrome containing operon in *Eggerthella lenta*, a common inhabitant of human gut, has been associated with the inactivation of digoxin, a cardiac glycoside (Saha *et al*., 1983). Thus, microbial members of human gut detoxify intestinal environment of human host.

2.3.3 Antimicrobial protection:

Mucus layer of human large intestine plays an important role in exclusion of pathogens (Johansson *et al*., 2008; Kim and Ho, 2010). However, mucosa of the small intestine is abrupt and discontinuous. Microbial members of the small intestine execute crucial protective activities against several detrimental pathogens. Beneficial microbes of gut environment induce the production of antimicrobial proteins (AMP) such as cathelicidins, C-type lectins, and defensins by paneth cells of the host by means of efficient pattern recognition receptor (PRR) operated mechanism (Hooper, 2009; Salzman *et al*., 2007). Microbial components like lipid A, peptidoglycan, flagella and bacterial RNA/DNA act as microbe-associated molecular patterns (MAMPs) that activate the PRRs (Carvalho *et al*., 2012; Cash *et al*., 2006). PRR-MAMP (pattern recognition receptor- microbe associated molecular patterns) interactions regulate activation of crucial signaling pathways associated with production of antimicrobial proteins and enhancement of mucosal barrier function. *B. thetaiotaomicron* and *Lactobacillus innocua* serve as key players in the production of antimicrobial proteins (Carvalho *et al*., 2012; Cash *et al*., 2006; Hooper *et al*., 2003). Lactic acid, produced as an end product of fermentation by *Lactobacillus*, hones antimicrobial efficacy of host lysozymes (Alakomi *et al*., 2000). Intestinal dendritic cells that regulate plasma cells to produce secretory IgA (sIgA) have been reported to be activated by members of the phylum *Bacteroides* (He *et al*.,

2007).

2.3.4 Immunomodulation:

Gut microbial members significantly modulate the immune system of human host in ensemble with innate and adaptive immune systems of the host. Gut associated lymphoid tissues (GALT), IgA producing plasma cells, effector and regulatory T cells, macrophages and dendritic cells actively participate in immunomodulatory process. Gut microflora has been found to produce Peyer's patches and isolated lymphoid follicles which eventually help in framing of GALT (Durkin *et al*., 1981). Bacterial masses of the human intestine have also been reported to enhance the normal functioning Foxp3+ T regulatory (Treg) cells. Polysaccharide A component of *Bacillus fragilis* has been associated with apt induction of Treg cells via TLR2 signaling pathway (Round *et al*., 2011). Short chain fatty acids (SCFA) like butyrate produced by the members of gut microbial community have been found to regulate the functions of Treg cells (Arpaia *et al*., 2013; Furusawa *et al*., 2013; Smith *et al*., 2013). Human gut-associated microbes have also been reported to influence the action of the innate lymphoid cells (ILCs) (Zelante *et al*., 2013).

2.3.5 Integrity of gut barrier:

Members of human gut microflora act as key components that control the structure and function of human gut. *B. thetaiotaomicron* regulate the expression profiles of small proline-rich protein 2A (sprr2A) associated with maintenance of desmosomes at the epithelial villus (Lutgendorff *et al*., 2008). Peptidoglycan components of microbial cell wall have been reported to influence TLR2 mediated signaling (Cario *et al*., 2007). Soluble proteins like p40 and p75 produced by *Lactobacillus rhamnosus* GG strain have been observed to avert cytokine induced apoptosis of the intestinal epithelial cells (Yan *et al*., 2011). Human gut microflora also finds its role in efficient development of gut structure. Microbial masses residing in human gut induce transcription factor angiogenin-3 that has been found to be associated with development of intestinal microvasculature (Stappenbeck *et al*., 2002).

2.4 Gut microbiota and related disorders:

Bacterial members residing in human intestinal environment crucially govern the 'steady state' of the gut. Proper balance between human immune

system and the commensal bacterial members hold the key of proper 'well-being'. Disruption of this equilibrium, also known as dysbiosis, results in several gastrointestinal disorders of major concern. Majority of the diseases involve inflammatory bowel diseases (IBD), irritable bowel syndrome (IBS), colorectal cancer (CRC) and gastric cancer. Disturbances in gut microbial balance have also been found to be associated with obesity and type 2 diabetes (Ley, 2010). *Helicobacter pylori* has been a notorious pathogen associated with gastritis, duodenal ulcers and gastric cancer and has generated a lot of concern (Neelapu and Pavani, 2013; Nammi *et al*., 2016). However, exact mechanisms underlying dysbiosis and gastrointestinal disorders still remain somewhat obscure. Various hypotheses have been put forward pertaining to the complex mechanisms. Some researchers have proposed the 'pathogen hypothesis' in which they have tried to correlate gastrointestinal diseases with infective potential of established and recognized pathogens. *Mycobacterium avium* subsp. *paratuberculosis,* an invasive pathogen associated with Crohn's disease, has been extensively studied (McFadden *et al*., 1987; Yoshimura *et al*., 1987). Another hypothesis has been proposed which states that disturbance in gut microbial composition and reduction in the pool of beneficial microbes provides a scope for the pathogenic members to thrive and proliferate (Pothoulakis, 1996). Some major health disorders associated with gut microbiota have been discussed below:

2.4.1 Inflammatory bowel disease:

Inflammatory bowel disease (IBD) includes ulcerative colitis (UC) and Crohn's disease (CD) and is marked by chronic inflammation of the gastrointestinal tract. UC and CD are chronic types of inflammatory bowel diseases that differ considerably in symptoms and inflammation signatures. CD, characterized by segmental inflammation of gut, has been suggested to be the upshot of complex interplay between host genetics and the gut microbial pool (Loftus, 2004; Elson, 2002). UC involves inflammation and formation of ulcers in colon. Dysbiosis of human gut results in the onset of these diseases (Lepage *et al.*, 2011; Martinez *et al.*, 2008).

It has been reported that patients suffering from UC and CD exhibit low population of *Roseburia*,

Phascolarctobacterium and certain other bacterial members representing families Enterobacteriaceae, Ruminococcaceae and Leuconostocaceae in their gut (Morgan *et al*., 2012). High frequency of *Clostridium* was also noted among the patients with UC and CD (Morgan *et al*., 2012). Analysis pertaining to low levels of *Faecalibacterium prausnitzii* in individuals affected with CD and UC (Frank *et al.*, 2007; Martinez-Medina *et al.*, 2006; Sokol *et al.*, 2009) have been instrumental in surmising a probable role of this microbe in preventing dysbiosis (Sokol *et al*., 2008). Pathogenic invasive *E. coli* has also been observed to be associated with CD (Darfeuille-Michaud *et al*., 2004).

2.4.2 Gastric cancer:

Gastric cancer is a worldwide problem and is one of the leading causes of deaths due to cancer (Lochhead and El-Omar, 2007). Gastric cancer is a global problem of tremendous concern and results in a huge loss of lives per year. *H. pylori*, a resident of human gut, has been found to be associated with manifestation of gastric cancer (Neelapu and Pavani, 2013; Nammi *et al*., 2016). However, only certain individuals bearing this pathogen develop gastric cancer which suggests that there might be other factors like host susceptibility, dietary habits and lifestyle that might contribute crucially in the onset of the disease. Rise in *H. pylori* population results in mild pangastritis (Smith *et al*., 2006). Prolonged colonization leads to significant alterations in intestinal physiology and gut microbial balance leading to gradual onset of cancer. Changes in gut microbial composition result in lower acid production in the gut which subsequently paves way for *H. pylori* pathogen to proliferate in gut niche. *H. pylori* has been reported to convert nitrite leading to the production of carcinogenic N-nitroso compounds in the intestinal environment (Blaser and Atherton, 2004).

2.4.3 Colorectal cancer:

Gut microbiota has been associated with establishment of colorectal cancer (CRC) in human beings (Arthur *et al.*, 2012; Kostic *et al.*, 2012; Plottel and Blaser, 2011). Members of the genus *Fusobacterium* have been reported to be linked with CRC manifestation (Castellarin *et al*., 2012; Kostic *et al*., 2012; McCoy *et al*., 2013). It has been revealed from 16S rDNA sequencing analysis and metagenomic analysis of

colorectal microbiome that the tumour tissues are enriched with several species of *Fusobacterium* which includes *Fusobacterium nucleatum, Fusobacterium mortiferum* and *Fusobacterium necrophorum* (Kostic *et al*., 2012; Castellarin *et al*., 2012). Furthermore, the tumour tissues have been found to display reduced population of Firmicutes and Bacteroidetes (Kostic *et al*., 2012; Castellarin *et al*., 2012).

2.4.4 Obesity and type 2 diabetes:

Obesity and related metabolic disorders are major health issues these days. Type 2 Diabetes (T2D) is also a grave problem that seeks careful medical attention. It has been reported from metagenomic studies of human gut microbiota that the microbial members play key roles in obesity (Ley, 2010; Ley *et al.*, 2005; Tilg and Kaser, 2011; Turnbaugh *et al.*, 2006). Interestingly, it has been observed that alteration in Firmicutes to Bacteroidetes ratio has been imperative in obesity and T2D (Ley *et al*., 2006). Sharp reduction in population of Bacteroidetes and increased frequency of Actinobacteria has also been demonstrated in obese individuals (Turnbaugh *et al*., 2009a).

2.5 Conventional methods to study human gut microbiota:

Proper assessment of gut microbial composition and elucidation of associated functions has been a challenging yet, interesting chore. Several techniques have been successfully implemented for gaining a better insight into the complex puzzles of human gut microbiome. Some commonly employed approaches have been discussed below:

2.5.1 Molecular approaches:

Rapid advancement of molecular approaches has prepared the platform for thorough investigations pertaining to human gut microbial composition. Proper analysis of microbial 16S ribosomal RNA gene enables extensive profiling of gut microbiome and assessment of phylogenetic relationships between different microorganisms constituting gut microbiota.

2.5.2 Microbial community fingerprinting:

Microbial community fingerprinting commonly refers to the collective molecular methods employed for extensive compositional analysis of microbial communities present in a sample of interest. Microbial community fingerprinting techniques

depend on multi-template PCR reactions for apt estimation of microbial community signatures or 'fingerprints'. These techniques have been extremely instrumental in analyzing microbial constitution of human gut. Common microbial community fingerprinting methods involve:

2.5.2.1 DGGE/TGGE:

Denaturing gradient gel electrophoresis (DGGE), or temperature gradient gel electrophoresis (TGGE) is one of the most popular microbial community fingerprinting techniques employed for accurate analysis of human gut microbial organization (Monstein *et al.*, 2000; Zijnge *et al.*, 2006). DGGE separates multi-template PCR products on the basis of genomic GC content of the organisms present in the sample of concern. DGGE has recently been used for estimation of gut microbial stability and subsequent correlation with health and disease of human host (Scanlan *et al.*, 2006; Zoetendal *et al.*, 1998).

2.5.2.2 T-RFLP:

Terminal-restriction fragment length polymorphism (T-RFLP) is another fingerprinting technique that finds applications in precise assessment of microbial community structure (Jernberg *et al.*, 2007; Kitts, 2001; Liu *et al.*, 1997; Wang *et al.*, 2008). T-RFLP is characterized by end labeling of PCR product with a fluorescent dye. Restriction signatures are produced by restriction digestion of the amplicons. However, terminal labeled fragment is only detected, based on its fluorescence, by automated sequencer.

2.5.2.3 Cloning and sequencing:

Comprehensive information is yielded from cloning and sequencing of 16S rRNA genes of bacterial members residing in human gut. Cloning based techniques of bacterial estimation tend to be accurate and precise. However, these methods involve huge cost and seem time consuming. Large-scale cloning endeavors pertaining to microbial composition analysis of various sites of human body have been effective in generating meaningful information (Aas *et al.*, 2005; Bik *et al.*, 2006; Eckburg *et al.*, 2005; Pei *et al.*, 2004).

2.5.3 FISH:

Fluorescent In Situ Hybridization (FISH) is a popular technique employed for direct estimation of bacterial masses present in a sample of interest. Fluorescently labeled oligonucleotide probes are used for bacterial enumeration. Epiflourescent microscopy, flow cytometry and

confocal microscopy allow detection of fluorescently labeled bacteria (Zoetendal *et al.*, 2004). FISH proves to be instrumental in inferring phylogenetic relationships. Proper performance largely depends on reliable sequence data set and expert analysis.

2.5.4 Quantitative PCR:

Quantitative PCR (qPCR) is an efficient approach for direct assessment of bacterial population in a given sample. Chemiluminscent or fluorescent reactions are employed for estimation of kinetic dynamics of product accumulation during PCR amplification for specific bacterial members. 16S rRNA gene is frequently used for the purpose.

2.5.5 G+C fractionation:

Percent guanine + cytosine (%G+C) fractionation performs assessment of complex microbial communities on the basis of genomic compositional variations. G+C fractionation is often used for estimation of microbial members of human gastrointestinal tract (Dicksved *et al.*, 2008). The concerned approach involves the separation of bacterial chromosomes in a density gradient based on the differences of genomic G+C content in presence of specific DNA-binding dyes (Apajalahti *et al.*, 1998). G+C fractionation method of bacterial estimation does not require primer and probe sequences and thus, seems advantageous over PCR and cloning based techniques of bacterial enumeration.

2.5.6 Multivariate data analysis:

Multivariate statistical tools and techniques seem inevitable in present day research. Gargantuan data resulting from molecular investigations of complex bacterial communities need to be processed effectively by means of statistical methods. Techniques like correspondence analysis (CA) (Edlund *et al.*, 2006), principal component analysis (PCA) (Jernberg *et al.*, 2005; Kaplan and Kitts, 2003; Wang *et al.*, 2004) and non-metric multidimensional scaling (nMDS) (Rees *et al.*, 2004) method are mostly used for extensive statistical profiling of complex data pertaining to human gut microbiome.

2.6 Bioinformatics as tool to explore the complexities of human gut microbiota:

2.6.1 Dawn of bioinformatics:

It was in the year 1953 when James Watson and Francis Crick proposed the twisted-ladder (double helix) structure of deoxyribonucleic acid (DNA). This

particular event forever changed the history of biological science and gave rise to modern molecular biology. In short order, their discovery yielded ground-breaking insights into the genetic code and protein synthesis. During the 1970s and 1980s, it helped to produce new and powerful scientific techniques, specifically recombinant DNA research, genetic engineering, rapid gene sequencing, etc. Around the same time, the term “bioinformatics” was coined by Ben Hesper and Paulien Hogeweg (Hesper and Hogeweg, 1970). Bioinformatics is by nature a cross-disciplinary field that began in the 1970s with the efforts of Margaret O. Dayhoff, Walter M. Fitch, Russell F. Doolittle and others and has gradually matured into a fully developed discipline. Initially, it was referred as ‘the study of information processes in biotic systems’ (Hesper and Hogeweg, 1970). However, its primary use since at least the late 1980s has been to describe the application of computer science and information sciences to the analyses of biological data, particularly in the areas of genomics involving large-scale DNA sequencing (Luscombe et al., 2001). The arrival of the INTERNET is another important milestone in the development of bioinformatics as a full -fledged discipline. This discipline represents the convergence of genomics, biotechnology and information technology, and encompasses analysis and interpretation of data, modeling of biological phenomena, and development of algorithms and statistics (Fenstermacher, 2005). The need for bioinformatics was further accelerated when the Human Genome Project (HGP) was launched in 1990. The aim of the project was to sequence the entire human genome. Information gleaned from the HGP is not very useful until the huge data is managed and interpreted in a proper way by the computational tools leading to the materialization of bioinformatics. The success of HGP opened the flood-gates for other genome sequencing projects. Gradually genome sequences of mouse, rat, worms, yeast and plants like rice, *Arabidopsis* were completed. The publication of huge amount of sequence data were greatly supported by development of high end computers, smart computational tools for large-scale annotation, functional classification of the proteins (Searls, 2000) and development of specific databases (Birney *et al*., 2002).

Availability of complete genome sequences of different organisms lead to the development of public repositories of gene data like GenBank (Benson *et al*., 2000), EMBL (Baker *et al*., 2000), DDBJ (Okayama *et al*., 1998), Protein DataBank (PDB) (Bernstein *et al*., 1977) and several others. After the formation of the databases, tools became available to execute various analyses. Two programs, which greatly facilitated the similarity search, were FASTA (Pearson and Lipman, 1988) and BLAST (Altschul *et al*., 1990). Many programs have been further developed since then. Accessibility of free and open source software has taken bioinformatics and its application to all -together new heights.

2.6.2 Bioinformatics based platforms suited for research pertaining to human gut microbiota:

Several 'omics' based research disciplines i.e., genomics (for genome data), proteomics (for protein data), transcriptomics (for gene transcription data), etc., have emerged with the enormous advancement of bioinformatics. The 'cause and effect' relationships of biological systems are being frequently employed by mathematical models and computational simulations for proper elucidation of biological complexities. Proper blend of knowledge-driven computational simulations and data-oriented bioinformatics holds the key in apt realization of the riddles of host-microbe interactions. Meaningful analysis of high-throughput sequencing data promises to unravel the enigma of the complex associations. Specific biological databases, providing access to annotated genomic data of bacterial populations, like IMG/M (Markowitz *et al*., 2014), SEED (Overbeek *et al*., 2005) and Greengenes (DeSantis *et al*., 2006) have opened up revenues to extract significant information. Annotation pipelines like RAST (Aziz *et al*., 2008) and MG-RAST (Meyer *et al*., 2008) have been instrumental for the wholesome purpose of characterizing bacterial masses. Systems Biology Markup Language (SBML) (Hucka *et al*., 2003) and BioModels repository (Le Novere *et al*., 2006) have been very useful for proper standardization of biochemical reaction models. Various bioinformatics based platforms suited for investigations pertaining to human gut microbiota have been described here:

2.6.2.1 Marker Gene Profiling:

Marker gene profiling (also known as gene amplicon sequencing) involves extraction of DNA from host-microbiota samples. DNA extraction is followed by amplification using primers specifically designed against ribosomal RNA genes. Initially, 454 pyrosequencing technology was pretty popular for the chore but nowadays, it has been replaced by Illumina sequencing which has been reported to produce equivalent results with a better coverage (Luo *et al*., 2012). QIIME (Navas-Molina *et al*., 2013) and MEGAN (Huson *et al*., 2011) have been efficient pipelines for analyzing data extracted via marker gene profiling. PICRUSt effectively predicts the functional potentials of gut associated microbes by analyzing the metagenome from the marker gene data (Langille *et al*., 2013). Marker gene profiling does not involve high cost and thus, aptly suits the purpose of large projects concerned with data extraction. Pathway enrichment followed by metabolic reconstruction has been suggested to be fruitful pertaining to proper analysis of metagenomic data (Abubucker *et al.*, 2012; Sharon *et al.*, 2011). Robust machine learning algorithms have been suggested to be useful for proper predictive analysis of data produced by marker gene profiling (Nakano *et al.*, 2014; Statnikov *et al.*, 2013). The major limitation of the technique is that it largely depends on pre-existing curated rRNA databases for taxonomic profiling which highlights the impotence of the method to successfully characterize novel bacterial species present in the samples (Dicksved, 2008).

2.6.2.2 Metagenomics:

Metagenomics is the branch of genomics that performs investigations by direct extraction and cloning of DNA from assembly of organisms (Handelsman, 2004). Comprehensive sequencing is necessary to pursue metagenomics based investigations. It is taken into presumption that human gut microbiome contains a large number of uncultured species and therefore, thorough sequencing is mandatory to proceed with metagenomics of gut samples (Emerson *et al.*, 2012; Hess *et al.*, 2011; Narasingarao *et al.*, 2012). Sophisticated genome assembly protocols (Bashir *et al.*, 2012; Goldberg *et al.*, 2006) and efficient single-cell sequencing technologies (Lasken, 2012; Shapiro *et al*., 2013)

have been suggested to aid proper metagenomic analysis. Due to extensive focus on human gut microbiome and availability of numerous bacterial genomes, assembly -free methods tend to be useful in thorough scrutiny of metagenomic samples (Carr *et al*., 2013; Luo and Moran, 2013). Rapid progress of next-generation sequencing (NGS) technologies has been a boost for the domain of metagenomics. NGS based techniques allow in-depth profiling and analysis of microbiome without the limitations of selection bias and constraints, associated with cultivation methods. NGS-based methods rely greatly on sophisticated bioinformatics based tools, regularly updated data repositories and functional know-how.

The main motif of gut-associated metagenomic analysis has been proficient characterization of and profiling of the microbial members framing the gut microflora. Similarity based alignment tools like BLAST serve the preliminary purpose of finding out regions of similarity between the raw or assembled sequences against a reference database and thus, provide an initial clue about the functionalities of the query samples. MG-RAST pipeline employs the M5nr database that is a large repository of non-redundant protein sequences from multiple sources (Weinstock, 2012; Wilke *et al*., 2012). QIIME pipeline facilitates taxonomic classifications. Functional and metabolic databases like KEGG Orthology (Kanehisa and Goto, 2000) or SEED subsystems (Overbeek *et al*., 2005) assist investigations pertaining to functionalities. Sophisticated statistical techniques are also frequently employed for extensive functional analysis (Kristiansson *et al*., 2009). Metabolic modeling tends to be useful in apt execution of function oriented studies (Abubucker *et al*., 2012; Jiao *et al*., 2013; Levy and Borenstein, 2013). Metagenomics has been extremely useful in unleashing the intricacies of human gut microbiome and has expanded the catalog of known genes.

2.6.2.3 Metatranscriptomics:

Metatranscriptomics involves extraction of RNA from samples of interest and is much more complex than DNA extraction technologies. The major concern of metatranscriptomics is to achieve good quality sequence and sufficient yield. Fast alignment tools like bowtie (Langmead *et al*., 2009) and SSAHA (Ning *et al*., 2001) find applications in proper alignment of

transcriptomic data, retrieved from the host-microbiota sample, with a set of representative bacterial genomes (McNulty *et al.*, 2011; Turnbaugh *et al.*, 2010; Xiong *et al.*, 2012). Simple BLAST has also been employed for the purpose (Gosalbes *et al*., 2011; Xiong *et al*., 2012) but the task requires extensive and rigorous computational support to align millions of sequence reads. Current pipelines use variations of this basic approach (Leimena *et al*., 2013; Xu *et al*., 2014). Metatranscriptomic data promise to be handy in exploring the intricacies of gut microbiota and provide vivid depiction of the host-microbiota interactome (Westermann *et al*., 2012; Xu *et al*., 2014). However, retrieval of sufficient bacterial RNA from combined host-microbiota sample has been a daunting task as host RNA prevails over bacterial RNA due to low biomass of the microbial members. Another crucial step is the removal of ribosomal RNA from bacterial RNA in order to avoid misinterpretations while analyzing metatranscriptomic data.

2.6.2.4 Metabolomics:

Metabolomes refer to the complete set of metabolites present in a cell/ organism or a community of organisms (Jordan *et al.*, 2009). Research pertaining to metabolome of human gut microbiota has been an exciting challenge these days. Metabolomes have been suggested to be crucial players of proper interactions and cross -talks between gut associated microbes and human host (Marcobal *et al.*, 2013; Nicholson and Lindon, 2008). Extensive profiling of metabolomes promise to confer large body of information regarding the complex interactions and also the health benefits and hazards associated with the concerned microbes of human gut (Larsen and Dai, 2015). (Marcobal *et al.*, 2013; Nicholson and Lindon, 2008).(Larsen and Dai, 2015).

2.6.2.5 Computational Modeling and Simulation:

Availability of complete genome sequences has made it feasible to execute computational and metabolic modeling of concerned microorganisms (Oberhardt *et al*., 2009; Thiele and Palsson, 2010). Huge data surging out of the various metagenomic projects also provide ample scope for apt simulation and metabolic modeling (Henry *et al*., 2010). Computational simulations do employ simplified techniques based on abstractions. Computational models not only prove fruitful for a single

organism but also can be employed efficiently for a community of organisms displaying significant interactions (Borenstein, 2012). This technique also finds its applications in supra-organism where there is significant assemblage of metabolic pathway associated genes reflecting all involved species (Borenstein, 2012).

2.6.3 Human Microbiome Project and MetaHIT:

The fact has been prominent that culture-independent approaches provide a better insight into the riddles of gut microbiota. High-throughput sequencing technologies (HTS) have been revolutionary in elucidating the complexities of human gut microbiome. Apart from Roche 454 pyrosequencing platforms, other high-throughput sequencing technologies, such as Illumina (San Diego, CA, USA), SOLid system (Applied Biosystems, Foster City, CA, USA), the Ion platforms (Life Technologies, Carlsbad, CA, USA) and SMRT system (Pacific Biosystems, Menlo Park, CA, USA) have become very popular recently (Caporaso *et al.*, 2011; Clarke *et al.*, 2009; Rosenstein *et al.*, 2012; Schadt *et al.*, 2010). The major limitation of 16S rRNA studies is that they do not provide any know-how pertaining to functional potential and viability of the concerned microbes existing in an ecosystem. Metagenomics based approaches tend to be more advantageous than 16S rRNA based studies because they efficiently characterize and profile the global genetic constitution of a community (Qin *et al.*, 2010; Turnbaugh *et al.*, 2009b). Furthermore, metagenomic analysis provides an insight into the functional attributes of the bacterial members of the concerned community (Kurokawa *et al.*, 2007; Qin *et al.*, 2010; Turnbaugh *et al.*, 2009b). Both the mentioned approaches require bioinformatics based expertise for meaningful inference (Kuczynski et al., 2011).

Human Microbiome Project (HMP) was launched in late 2004 by National Institute of Health, USA, with a motif to assess the baseline constitution of human ‘superorganism’. High levels of scientific expertise and adequate funds were employed to cater the need. The main purpose was to execute an extensive profiling of microbes dwelling in various body sites of human beings that included various parts like gut, mouth, skin, uro-genital tracts, etc. Various detection schemes like whole genome shotgun sequencing

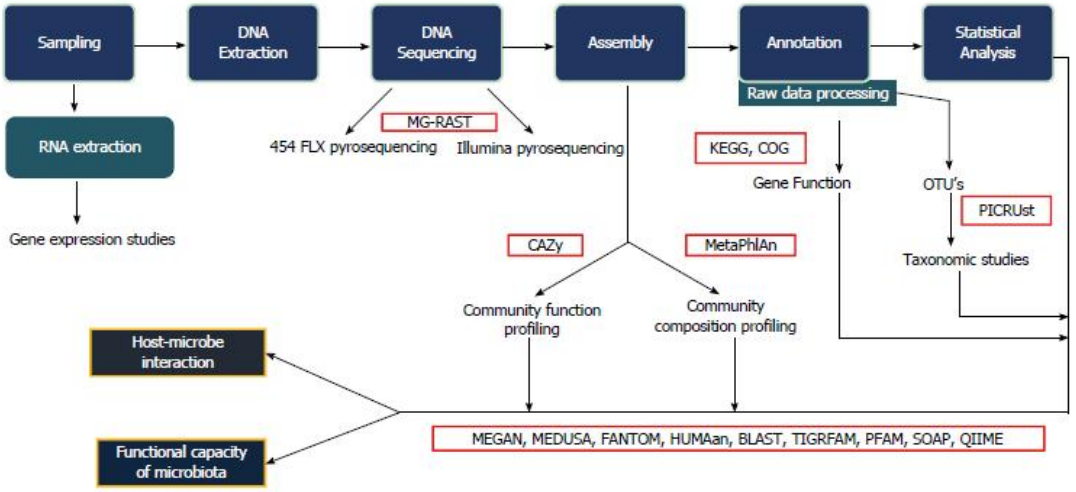

Figure 2.1: Recently employed bioinformatics workflow in study of human gut microbiome (reproduced with permission from: Jandhyala *et al.* (2015). *World J Gastroenterol* **21**: 8787-8803) - This figure explains the various steps involved in the bioinformatics analysis, starting from collection of samples, extraction, sequencing and statistical analysis. The interaction between host and microbes along with the functional capacity of the microbiota can be studied. MG-RAST: Metagenomics rapid annotation using subsystem technology; CAZy: Carbohydrate active-enzymes; MetaPhlAn: Metagenomic phylogenetic analysis; KEGG: Kyoto encyclopaedia for genes and genomes; COG: Clusters of orthologous group; PICRUst: Phylogenetic investigation of communities by reconstruction of unobserved states; MEGAN: Meta genome analyzer; MEDUSA: Metagenomic data utilization and analysis; FANTOM: Functional annotation and taxonomic analysis of metagenomes; HUMAan: Human microbiome project unified metabolic analysis network; BLAST: Basic local alignment search tool; TIGRFAM: Protein sequence classification; PFAM: Protein families; SOAP: Short oligonucleotide analysis package; QIIME: Quantitative insights into microbial ecology

technologies and metagenomic analysis were opted for a thorough characterization of the commensal residents of human body. HMP has been a fascinating initiative aimed at identification and characterization of diverse microbial communities resident in various sites of healthy human host including gastrointestinal tract and promises to reveal the probable connections of gut microbiota and gastrointestinal disorders (Arumugam *et al.*, 2011; Qin *et al.*, 2010).

The European Metagenomics of the human intestinal tract (MetaHIT) (Arumugam *et al.*, 2011; Qin *et al.*, 2010) is another mammoth sequencing endeavor that has been addressed for proper realization of the baseline composition of healthy gut microbiota and how it differs from diseased state. The main targeted diseases are obesity and inflammatory bowel diseases (IBD) (Qin *et al.*, 2010). MetaHIT particularly deals with large-scale high-throughput sequenced data of healthy and diseased gut microbiome and compares them in order to excavate the compositional variations (Qin *et al.*, 2010). MetaHIT has been extremely instrumental in correlating gut microbial constitution and human intestinal pathologies. Bioinformatics based workflow employed for research

pertaining to human gut microbiome has been illustrated in Figure 2.1 which has been reproduced from the review work performed by Jandhyala and colleagues (Jandhyala *et al*., 2015), with permission.

2.6.4 Recent trends in bioinformatics of human gut microflora:

In the post genomic era, application of bioinformatics based tools in comparative genomics has led to the belief that every genome has its own story. Particularly the genetic code and its usage preferences is one of the most interesting aspects of biological science. Proper codon and amino usage analysis promises to unravel an extensive body of information towards the designing of an accurate codon optimization methodology.

Variations in codon usage have been attributed to various determinants such as compositional constraints (Hou and Yang, 2002; Karlin and Mrazek, 1996), selection pressure for translational efficiency (Romero *et al*., 2003), gene expression level (Duret and Mouchiroud, 1999; Romero *et al.*, 2003; Sharp and Li, 1986), abundance of transfer RNA (tRNA) (Duret, 2000; Ohkubo *et al.*, 1987), replicational-transcriptional selection (Das *et al.*, 2005; Guo and Yu, 2007; Guo and Yuan, 2009; McInerney, 1998; Romero *et al.*, 2000), codon-anticodon interactions (Shi *et al*., 2001), protein secondary structure (Gu *et al*., 2004), gene length (Moriyama and Powell, 1997), hydropathy of proteins (Romero *et al*., 2000), stability of mRNA folding (Kahali *et al*., 2007), etc.

Compositional constraints and natural selection for translation are reported to be the major factors that crucially govern the codon usage patterns of various prokaryotic as well as eukaryotic genomes (Gu *et al.*, 2004; Romero *et al.*, 2003; Sharp *et al.*, 1993). Our group previously reported such an instance in GC rich Actinobacteria, *Frankia*, where both GC compositional constraint and natural selection for translation influenced the codon usage trends (Sen *et al*., 2008). However, some extremely AT/GC rich unicellular organisms have been found to display codon usage signatures solely influenced by base compositional bias (Andersson and Sharp, 1996; Ohama *et al.*, 1990). Cases have also been reported where translational selection factor has been found to govern the codon usage in highly expressed genes of an organism. Such occurrences have been interpreted to be outcomes of preferential usage of

specific sets codons, which are recognized by the most abundant isoacceptor tRNAs (Ikemura, 1981; Ikemura, 1982). Organisms like *E. coli* (Ikemura, 1981), *Drosophila melanogaster* (Moriyama and Powell, 1997) and *Caenorhabditis elegans* (Duret, 2000) show trends where codon usage patterns of highly expressed genes are dictated by the abundance of tRNAs. Usage of preferred codons, with corresponding abundant tRNA population, have been inferred to be accountable for three to six fold differences in rates of translation (Robinson *et al*., 1984) and ten-fold differences in translational efficacy (Precup and Parker, 1987). Similarly, amino acid usage variations in microbial genomes have also been observed to be governed by several factors like hydrophobicity, aromaticity and expression level of the respective gene products (Das *et al.*, 2006; Lobry and Gautier, 1994). Thus, several crucial factors have been found to be accountable for discrepancies in codon and amino acid usage behavior.

Biased usage of single codons is not the sole factor that significantly dictates translational efficiency. Recently, it has been suggested that predisposition in the arrangement patterns of successive synonymous codon pairs considerably administer translational accuracy (Buchan *et al.*, 2006; Fedorov *et al.*, 2002; Gutman and Hatfield, 1989; Irwin *et al.*, 1995). Pairs of synonymous codons that appear in successive fashion, regardless of the number of codons that code for different amino acids (non-synonymous codons) occurring between them, are termed as successive synonymous codon pairs (Cannarozzi *et al*., 2010). It has been reported in prokaryotic and eukaryotic forms of life that the usage of synonymous codon pairs do not occur by chance and such an usage clearly has a motive to optimize the speed and fidelity of protein synthesis (Cannarozzi *et al.*, 2010; Fredrick and Ibba, 2010; Guo *et al.*, 2012; Plotkin and Kudla, 2011; Zhang *et al.*, 2013). Biased usage of identical pairs of codons has been well attributed to maximize translational efficacy in bacterial, archaeal and eukaryotic forms of life (Cannarozzi *et al*., 2010; Guo *et al*., 2012; Zhang *et al*., 2013). Non-identical codons pairs that are accepted by isoaccepting tRNAs (non-identical co-tRNA codon pairs) have also been observed to be mostly over-represented while codon pairs read by non-isoaccepting tRNAs

(tRNAs without isoacceptors) appear to be frequently under-represented. Thus, preferential usage of synonymous codon pairs and 'codon autocorrelation' (also termed as 'codon reuse') provide newer insight to properly elucidate translational mechanisms.

Completely sequenced bacterial genomes representing human gut microflora have provided a firm pedestal and scope for meticulous investigation of the genomic and proteomic traits and proper elucidation of the interaction patterns employed by these organisms for successful adaptation in the human gut. Codon and amino acid usage properties of certain bacterial members residing in human gut have already been explored.

AT compositional constraint has been found to be the major driving force in shaping codon usage patterns in *Helicobacter pylori* (Lafay *et al*., 2000) and the genus *Eubacterium* (Shende *et al*., 2013). Recently, Nayak (2012), reported that factors like gene expressivity, mutational bias and gene length have been instrumental in shaping codon usage variations in the genus *Lactobacillus*. Amino acid usage was found to be driven by hydrophobicity and aromaticity of the encoded protein products in the same genus (Nayak, 2012). However, thorough understanding of the factors underlying the complex codon and amino acid usage behavior in various crucial genera like *Bifidobacterium*, *Ruminococcus*, *Bacteroides*, *Butyrovibrio*, *Enterococcus*, *Coprococcus*, *Fusobacterium*, *Roseburia*, etc., is yet to be accomplished. Therefore, in this present work we intend to explore the puzzles of codon and amino acid usage patterns among the bacterial members of the human gut and simultaneously compare the observed patterns with that of human host. Such a comparison might provide a vivid portrait about the adaptive strategies employed by the groups of interesting microbes for proper abode in human gut.

Ever since the inception of Human Microbiome Project and the rapid progress in sequencing of the major bacterial members of human gut, enormous genomic and proteomic data have been generated. Comparative genomics and proteomics based research have been significantly effective in inferring in-depth idea about various genomic and proteomic traits and signatures of the gut microbes.

Detailed comparative genomics and proteomics based analysis of various microbial members of human gut have already been accomplished. Comparative investigations of genomic and proteomic data in various probiotic genera like *Bifidobacterium* and *Lactobacillus* have been effective in elucidating the symbiotic relationship of the microbes with human host (Lukjancenko *et al*., 2012). *E. coli*, an important human gut resident, has also been well explored on the basis of genomic and proteomic features (Lukjancenko *et al*., 2010). Comparative genomic analysis of commensal and pathogenic isolates of *E. coli* revealed that the *E. coli* pangenome consists of a repertoire of more than 13,000 genes which might be associated with pathogenic manifestations even in the commensal isolates (Rasko *et al*., 2008). Multi-omics system analysis of *E. coli* B and *E. coli* K-12 strains have provided a clear insight into cellular physiology and metabolism (Yoon *et al*., 2012). *Fusobacterium* is another dominant genus of human gastrointestinal tract which has been reported to be associated with a wide spectrum of infections and diseases (Castellarin *et al*., 2012). Genetic elements in *Fusobacterium nucleatum* subsp. *vincentii* and *Fusobacterium nucleatum* ATCC 25586 have been inferred to confer resistance against various antibiotics like acriflavin, bacitracin and bleomycin (Kapatral *et al*., 2003). Furthermore, phylogenetic analysis has also been instrumental in deciphering the evolutionary lineage of *F. nucleatum* (Mira *et al*., 2004). Recently, an online *Fusobacterium* comparative genomic analysis platform - FusoBase has also been developed which provides a comprehensive platform for comparative genomic and proteomic scrutiny of several species of *Fusobacterium* (Ang *et al*., 2014). Comparative genomic investigations of various strains of *H. pylori* have revealed the characteristic features of various pathogenicity islands and associated infective mechanisms exhibited by the microbes (Alm *et al.*, 2000; Kumar *et al.*, 2015; Schott *et al.*, 2011). Various enterococcal members like *Enterococcus faecalis*, *Enterococcus faecium*, *Enterococcus casseliflavus* and *Enterococcus gallinarum* have been studied well from comparative genomics based perspective and such investigations have produced meaningful information pertaining to their adaptive and

pathogenic policies in the human intestinal niche (Palmer *et al*., 2012). Acclimatization tactics employed by the genus *Veillonella* has also been elucidated with the help of comparative genomics and proteomics based scrutiny (Vesth *et al*., 2013).

Carbohydrate-Active enZymes (CAZymes) have been identified and characterized in various bacterial genera like *Ruminococcus*, *Bacteroides*, *Bifidobacterium*, *Veillonella* and *Eubacterium* (Bottacini *et al.*, 2012; Christopherson *et al.*, 2014; El Kaoutari *et al.*, 2013; Flint *et al.*, 2012; White *et al.*, 2014). Among the myriad of genes that have been identified in the human gut microbiome, CAZymes are of special interest as these enzymes are inevitably required for digesting complex dietary polysaccharides. Multi-enzyme cellulosome components, associated with efficient degradation of cellulose, have also been thoroughly identified in various members of the genus *Ruminococcus* (Ben David *et al.*, 2015; Brulc *et al.*, 2011). Many crucial bacterial genera like *Bacteroides*, *Ruminococcus*, *Eubacterium*, *Coprococcus*, *Faecalibacterium*, etc., residing in the human intestine still remain unexplored from the aspect of comparative genomics and proteomics based know-how. Accordingly, in this present approach we plan to execute extensive genomic and proteomic profiling of certain vital bacterial members of human gut to gain a deeper insight into the mechanisms of successful residence in the human intestinal niche.

Bacteria communicate and interact among themselves and with their concerned hosts by means of sophisticated protein secretion machinery. The complete set of secreted proteins in any organism is often referred to as secretome (Ranganathan and Garg, 2009). Secretomes have been reported to facilitate cellular cross-talks, interaction, communication and cell migration and thus, seem inevitable for survival of any organism (Tjalsma *et al.*, 2004).

Bacterial members of human gut efficiently interact with host cells and execute several vital functions that not only assure their existence but also benefit human host. Research pertaining to secretome characterization and profiling has been a topic of enormous thrill. Considerable progress has been achieved in analyzing secretome

behavior of the genus *Lactobacillus*. Sets of secretory proteins have been identified in *Lactobacillus plantarum* WCFS1 and have been reported to efficiently interact with host environment (Boekhorst *et al.*, 2006; Mathiesen *et al.*, 2009). Extensive profiling of secretomes in *L. plantarum* has also been accomplished successfully (Mathiesen *et al.*, 2008). Recently, a database pertaining to secretomes in *Lactobacillus* has been created and contains a large repertoire of predicted and experimentally validated sets of secretory proteins from various strains of *Lactobacillus* (Zhou *et al.*, 2010). Thorough investigations of the extracellular secretory proteins in certain probiotic members of *Lactobacillus*, *Bifidobacterium* and *Escherichia* have been instrumental in deciphering the interaction mechanisms with mucosal cells of human host (Sanchez *et al.*, 2010). In spite of considerable progress in the concerned field, large-scale identification and investigation of secretomes of human gut-associated bacterial members remains somewhat unexplored. Therefore, in the present study we plan extensive characterization of the secretomes and intend to unravel the functional and evolutionary complexities of the crucial cellular components.

Human beings have been vulnerable to various bacterial infections right from antiquity. Rapid pace of genome sequencing technologies and progress in bioinformatics and cheminformatics based research fields have offered considerable prospects in the domain of drug discovery. Availability of completely sequenced genomes of both host and pathogens has built the pedestal to successfully design drugs against menacing pathogens employing subtractive genomics based drug target identification in the concerned agents of infection (Allsop, 1998; Stumm *et al*., 2002).

Bacterial residents of human gut severely regulate the ‘steady state’ of the host. Proper balance in gut microbial framework seems to be a crucial determinant associated with the ‘well-being’ of human system. Alterations in gut microbial composition pronounce severe impact on human health. Besides harboring the beneficial microbes, human beings also quay some pathogenic microbes that cause serious infections like inflammatory bowel diseases (IBD), irritable bowel syndrome (IBS), colorectal cancer (CRC) and gastric

cancer.

H. pylori is a detrimental pathogen that resides in human gut. It has been found to be associated with serious consequences like pangastritis (Smith *et al*., 2006), gastric and duodenal ulcers and gastric cancer (Neelapu and Pavani, 2013; Nammi *et al*., 2016). Due to its devastating effects, *H. pylori* has gained a lot of medical attention. Proper drug target identification in various *H. pylori* strains like HpB38, HpP12, HpG27, Hpshi470, and HpSJM180 has been successfully accomplished (Neelapu and Pavani, 2013). Characterization and structural profiling of novel potential drug targets has also been executed in *H. pylori* HPAG1 strain (Sarkar *et al*., 2012). However, screening and identification of putative drug targets in *H. pylori* 35A strain, a pathogenic member of human gut, still remains to be explored. Recently, Ali and colleagues (Ali *et al.*, 2015) executed comparative genomics based analysis to predict conserved therapeutic targets in 39 strains of *H. pylori* which included *H. pylori* 35A strain and revealed certain potential targets. However, a complete hunt and characterization of potential drug targets, aided by molecular docking investigations, remains unaccomplished. Accordingly, in this present approach we intend to perform comprehensive profiling and screening of tentative drug targets in *H. pylori* 35A and subsequent validation by molecular docking studies.