Chapter 1
Introduction

Human beings, similar to other higher multicellular organisms, live in close concord with numerous microbes that co-evolve and co-exist with them (Backhed *et al*., 2005). Microorganisms that reside on and inside human beings are collectively referred to as human microbiota or human microflora (Kunz *et al*., 2009). Bacterial population of human microbiota outnumbers the human somatic cells by tenfold and thus, human beings can be rightfully considered as heavily colonized 'microbial depots' (Ley *et al.*, 2006a; Savage, 1977). Humans are sterile during gestation and gradually acquire microbes during birth while passing through the birth canal. Prokaryotic and eukaryotic organisms, constituting human microbiome, bear a large repertoire of protein encoding genes that by far exceeds the gene pool found in the human genome (Backhed *et al*., 2005; Qin *et al*., 2010). Microbes extensively colonize the gastrointestinal, respiratory and genitourinary tracts and the skin surface of humans (Chiller *et al*., 2001; Hull and Chow, 2007; Neish, 2009; Verstraelen, 2008). Gastrointestinal tract (GIT) has been reported to be the most heavily colonized site of the human body and colon itself harbors more than 70% of all the microorganisms that make up human microbiota (Ley *et al*., 2006a; Whitman *et al*., 1998). Large surface area of the human gut (approximately 200-300m^2) (Gebbers and Laissue, 1989) makes it an ideal niche for the microbes to thrive and vibe. Furthermore, human GIT serves as a rich source of various nutrients that are utilized by the microbes for survival and thus, promises to be a preferred site for colonization (Sun and Chang, 2014). Colonization of human gut initiates right from birth when there is passage through the birth canal and subsequent exposure to various microbes. Human gut microbiome

tends to be relatively simple during the first year of its establishment (Mackie *et al*., 1999; Mandar and Mikelsaar, 1996). However, after the first year, composition of gut microbiome gradually starts to mature and stabilize finally resembling that of a healthy young adult (Mackie *et al*., 1999; Mandar and Mikelsaar, 1996). Human gut microflora, on full maturity, acts as an 'essential' acquired organ and executes several vital functions associated with the 'well-being' of human host (Karlsson, 2014).

Microbial composition of human gut is heterogeneous. Majority of the microbes that dwell in human gut represent strict anaerobes (Gordon and Dubos, 1970; Harris *et al.*, 1976; Savage, 1970). However, few aerobic bacteria have also been reported to be present in human gastrointestinal tract (Gordon and Dubos, 1970; Harris *et al.*, 1976; Savage, 1970). Firmicutes and Bacteroidetes are the major bacterial phyla that occur in human gut. Certain members representing Actinobacteria, Proteobacteria, Verrucomicrobia and Fusobacteria also display signs of existence, but in lower proportions (Eckburg *et al*., 2005; Tremaroli and Backhed, 2012). Members of the genera *Enterobacter* and *Enterococcus* initiate colonization of human GIT. Subsequently, members representing various other bacterial genera like *Bifidobacterium*, *Clostridium* and *Bacteroides* start inhabiting human intestinal environment (Adlerberth and Wold, 2009). A comprehensive list of the potential microbes that are encountered in healthy human gastrointestinal tract has been provided in Figure 1.1 which

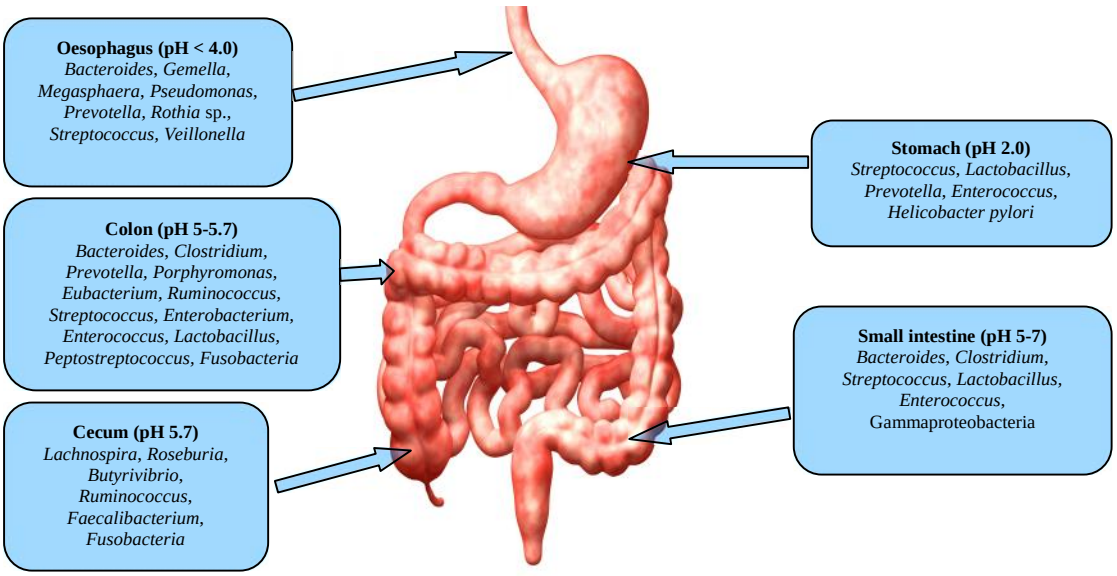

Figure 1.1: List of bacterial members that reside in various parts of human gastrointestinal tract (modified from: Jandhyala *et al*. (2015). *World J Gastroenterol* **21**: 8787-8803)

has been reproduced from the review work performed by Jandhyala and colleagues (Jandhyala *et al*., 2015), with permission.

The most diverse and abundant bacterial phylum that resides in human intestine is Firmicutes. Members of the concerned phylum constitute more than 50% of gut microbial population (Rajilic-Stojanovic and de Vos, 2014). Gastrointestinal Firmicutes mainly comprise of four classes: Bacilli, Clostridia, Erysipelotrichi and Negativicutes. The class Bacilli includes several bacterial genera like *Lactobacillus*, *Enterococcus*, and *Streptococcus* that dominate the upper gastrointestinal tract (Rajilic-Stojanovic and de Vos, 2014). *Lactobacillus*, an important genus of the phylum Firmicutes, dwells mostly in the small intestine of humans (Reuter, 2001) and has been reported to produce lactic acid as end product of homo- or hetro-fermentative metabolism (O'Sullivan *et al*., 2009). Various members of the genus *Lactobacillus* exert crucial probiotic functions associated with intestinal homeostasis (van Baarlen *et al*., 2009). *Lactobacillus acidophilus* NCFM has been reported to encode a number of permeases, glycolases and peptidases for rapid uptake and utilization of sugars and amino acids from human intestine, especially the upper gastrointestinal tract (Altermann *et al*., 2005). *L. acidophilus* NCFM also bears a number of cell-surface proteins, such as mucus- and fibronectin binding proteins that facilitate proper adherence to the intestinal epithelium and allow exchange of signals with the host intestinal immune system (Altermann *et al*., 2005). *Lactobacillus plantarum* WCFS1 has been reported to contain a large number of genes related to carbohydrate transport, utilization and production of exopolysaccharides (Kleerebezem *et al*., 2003). *Lactobacillus rhamnosus* ATCC 53013 has been reported to be a vital probiotic organism which facilitates amelioration of human intestinal environment (Ljungh and Wadstrom, 2006)). *Lactobacillus salivarius* UCC118 has been found to be associated with the production of antimicrobial peptides and bacteriocins and thus, aid pathogen exclusion from human intestine (Corr *et al*., 2007).

The genus *Ruminococcus*, representing the family Ruminococcaceae and phylum Firmicutes, is another established member of the human gastrointestinal tract. Cellulolytic

ruminococcal species like *Ruminococcus albus* 7 and *Ruminococcus champanellensis* 18P13 have been reported to degrade undigested cellulose components of human diet (Christopherson *et al*., 2014). Amylolytic member *Ruminococcus bromii* L2-63 efficiently degrades dietary starch that remains non-degraded in human intestine (Ze *et al*., 2012). Thus, various species of *Ruminococcus* enhance proper digestion of human host.

Eubacterium is another crucial genus of the family Eubacteriaceae, reflecting class Clostridia, that plays significant role in maintaining healthy state of human gut. Members of the genus *Eubacterium* produce butyrate that acts as a source of energy for the epithelial cells of human gut and also exhibits anti-cancer and anti-inflammatory activities (Hamer et al., 2008).

Members of the genus *Bacteroides*, reflecting phylum Bacteroidetes, are found to be prevalent within human GIT. Various species of *Bacteroides* utilize a wide range of polysaccharides and thus, enhance human digestion by breakdown of complex undigested carbohydrates (Martens *et al*., 2009). However, certain species of *Bacteroides*, namely *Bacteroides fragilis* has been associated with pathogenesis in human host (Wexler, 2007).

Actinobacteria is a major bacterial phylum present in human gastrointestinal environment. The genus *Bifidobacterium*, representing phylum Actinobacteria, is one of the most crucial genera within the human gut (Mitsuoka, 1990). Several bifidobacterial species have been reported to render crucial beneficiary effects on the physiological 'well-being' of human host (Ventura *et al*., 2007). *Bifidobacterium* produces organic acids like lactic acids, acetic acids and some bioactive metabolites as end products of fermentation. These metabolites render strong bactericidal effects, thus, inhibiting the pathogenic microbes to proliferate in the host intestinal niche (Ishibashi *et al*., 1997; Lau and Liong, 2014; Wei *et al*., 2012). Several species of the genus *Bifidobacterium* ameliorate intestinal environment by an appropriate degradation of the putrefactive products of various metabolic pathways (Ishibashi *et al*., 1997) and by efficient utilization of non-digestible oligosaccharides (Wei *et al*., 2012). Various strains are also believed to confer immunomodulatory effects in

bolstering the host intestinal defense against pathogenic microbes by regulating the production of specific and nonspecific antibodies, by their antitumor effects and by offering resistance against a broad range of bacterial toxins (Yamazaki *et al*., 1991).

Bacterial members like *Escherichia coli* and *Helicobacter pylori*, belonging to the phylum Proteobacteria, are also found in human intestine in considerable proportions (Rajilic-Stojanovic and de Vos, 2014). Various strains of *E. coli* display diverse functional activities that range from beneficial probiosis (Kruis *et al*., 2004) to certain pathogenic consequences like diarrhea (Ron, 2006). *H. pylori*, residing in human gut, has been associated with several threatening diseases that include gastric and duodenal ulcers (Fock *et al*., 2013) and gastric adenocarcinoma (Graham, 2000).

The genus *Fusobacterium* (representing phylum Fusobacteria) is frequently encountered in human GIT. Members of the genus have been reported to be associated with severe diseases like ulcerative colitis, colonic inflammation and colorectal cancer (Rajilic-Stojanovic *et al*., 2013; Castellarin *et al*., 2012).

Microorganisms of healthy human gut are known to bestow a series of health benefits on concerned human host. The beneficial effects include protection against pathogens, enhancement of host nutrition, metabolism and immune modulation (O'Hara and Shanahan, 2006; Sekirov *et al*., 2010). Intestinal microbes also exhibit anti-inflammatory and anti-tumor effects that help to maintain colonic equilibrium (Guinane and Cotter, 2013). However, perturbance in gut microbial community have been reported to cause inflammatory bowel disease and ulcerative colitis (Frank *et al*., 2007). Disruption of gut microbial balance, commonly referred to as dysbiosis, has been associated with various diseases like obesity (Ley *et al*., 2006b; Zhang *et al*., 2009) and diabetes (Qin *et al*., 2012). Therefore, it would be logical to consider human beings as ‘superorganisms’ colonized heavily by microbes that exert both beneficial and ill effects on human health.

In the era of genomics, biological data are being produced at a phenomenal rate. The first complete genomic sequence obtained was that of *Haemophilus influenza* (Fleischmann

et al., 1995). Since this major achievement, many genomes have been sequenced and analyzed. These include members of three life domains: Bacteria, Archea and Eukarya. The first genome of a mammalian lactic-acid bacterium, that of *Lactococcus lactis*, a microorganism of great industrial interest, was completed in 2001 (Bolotin *et al.*, 2001) More recently, the genomes of numerous other lactic-acid bacteria (Klaenhammer *et al.*, 2002), bifidobacteria (Kim *et al.*, 2009) and other intestinal microorganisms (Siezen *et al.* 2012*;* Xu *et al.*, 2003; Petrof *et al.*, 2004) have been sequenced. The major driving force for the development of genomics has been the completion of Human Genome Project in 2003. As a result of this surge in data, computers have become indispensable tools in biological research. Such an approach is ideal because of the ease with which computers can handle large quantities of data and probe the complex dynamics observed in nature. Thus, a new discipline, Bioinformatics, has emerged. Bioinformatics is defined as the application of computational techniques to understand and organize the information associated with biological systems and macromolecules. Bioinformatics in the broadest sense includes research in the domains of genome composition, genome expression, proteome analysis, genome and proteome engineering (Perez-Iratxeta *et al.*, 2007). National Institute of Health (NIH), USA, realized the necessity of studying the complexities of human-associated gut microbes at a greater depth and horizon and accordingly, launched the Human Microbiome Project (http:// www.hmpdacc.org/) (HMP). Gargantuan data produced from HMP has provided ample revenues to explore the complex interactions between human host and associated gut microbes. Emergence of bioinformatics has opened a new window for providing novel understanding into the adaptive policies employed by the bacterial members of the human gut. Comparative genomics, together with functional studies, has led to significant advances in this field over the past decade. In the era of genomics, it has been feasible to identify the crucial genetic elements that establish the platform for successful residence of the bacterial populations in the human intestinal environment. Metagenomics is the branch of genomics that performs

investigations by direct extraction and cloning of DNA from collection of organisms (Handelsman, 2004). Apart from HMP, Metagenomics of the Human Intestinal Tract (MetaHIT) has been another large-scale sequencing project that has offered scopes to address the complexities of human gut-associated microbial flora. Genomics coupled with proteomics and transcriptomics are now important high throughput techniques for qualifying and analyzing both gene and protein expression, discovering new gene and protein products, and perhaps these techniques hold the key for revealing the hidden facets of host-microbe interactions.

Advancement of genome sequencing technologies has revolutionized biological research. Plethora of information regarding genome profile can now be mined with the availability of fully-sequenced genomes. Study of codon usage patterns of several genes and genomes is a popular technique to characterize and analyze genomic trends from a bioinformatics-based perspective. Codon usage patterns and preferences vary significantly within and between organisms (Grantham *et al.*, 1981; Sharp *et al.*, 1988; Zhou and Li, 2009). The phenomenon of differential codon usage was proposed by Grantham and colleagues (Grantham *et al.*, 1981) in the 'genome hypothesis' theory and it was also stated that codon biases are usually species specific. Codon and amino acid usage profiling of several prokaryotic and eukaryotic forms of life have been successfully accomplished as of now.

Many indices have been proposed to properly elucidate the factors underlying the complex patterns of codon usage and to measure the degree and direction of codon bias (Sharp and Li, 1987). Studies of codon usage can be performed using parameters like GC (Guanine and Cytosine) content of the concerned organism, GC3 (Guanine and Cytosine at the third synonymous codon position) content, relative synonymous codon usage (RSCU) (Peden, 1999), effective number of codons (Nc) (Wright, 1990) and frequency of optimal codons (Fop) (Ikemura, 1981). Codon adaptation index (CAI) has been proposed as an efficient tool in analyzing the patterns of codon usage within a gene relative to a reference set of genes (usually ribosomal protein genes) (Sharp and Li, 1987). This index has been shown to correlate significantly with mRNA expression levels (Ikemura, 1981) and

has been used to predict sets of highly expressed genes in various organisms (Sharp and Li, 1987; dos Reis *et al.*, 2003; Martin-Galiano *et al.*, 2004; Wu *et al.*, 2005).

Availability of completely sequenced bacterial genomes representing human gut microbiome and the progress of the Human Microbiome Project has now made it possible to explore the complex riddles of codon and amino acid usage of these organisms and properly address the interplay of various factors that contribute to varying traits.

The rapid progress of genome sequencing has opened the flood-gates to explore gamut of data pertaining to genome dynamics and genome complexities in both prokaryotic and eukaryotic life forms. Comparative genomics based analysis attempts to take advantage of the information provided by the signatures of selection to understand the function and evolutionary processes that act on genomes (Lukjancenko *et al.*, 2012). Comparative genomics based approaches promise to reveal genetic variation among concerned sets of genomes. Strategic grouping of the genes into functional groups or families has been a smart way of extracting meaningful information (Zakham *et al*. 2012). Such grouping scheme is based on protein sequence similarity, as this approximately predicts conservation of gene function.

Systematic characterization and profiling of the core genome and pan-genome of organisms of interest has been an efficient technique to elucidate the puzzles of speciation and genomic variations. Core genome is referred to as the conserved pool of genes shared by the strains of a concerned species of interest (Medini *et al.*, 2005). Dispensable genome reflects the pool of genes present in some strains of a species but absent in others (Medini *et al.*, 2005). Pan-genome is commonly referred to as the collective repertoire of the core genes, the dispensable set of genes and the unique genes present only in a single strain (Rouli *et al*. 2015). The essence of a species, in terms of its fundamental biological processes and derived traits from a common ancestor, is linked to the core genome.

Rapid progress of HMP and MetaHIT has opened the revenues to explore the features and interaction patterns of inherent bacterial members of human gut. Comparative investigation of the bacterial genomes representing human

gut microflora, aided by a comprehensive profiling of the pan-genome, core genome and the unique gene sets, promise to bestow extensive information pertaining to the adaptive policies employed for successful residence in human gut.

Protein secretion in bacteria plays an imperative role in communication and cross-talk with other bacterial communities and also with host niche. Exploring the intricacies of bacterial communication and signaling with surrounding host environment has been a challenging chore of present day biological research. Secretomes have been defined as the complete set of proteins secreted by a cell (Ranganathan and Garg, 2009) and are associated with a broad range of functions and critical biological processes, such as cell-to-cell communication and cross-talks, cell migration, and most inevitably virulence and potential infective strategies in disease mechanism (Tjalsma *et al.*, 2004). The signal peptide part of the secreted protein, which is generally composed of around thirty amino acid residues, transports the newly synthesized protein to the protein-conducting SecE and SecY channels associated with the plasma membrane (Leversen *et al.*, 2009). Signal peptides in most cases are reported to possess three domains: a positively charged n-terminus (n-region), a stretch of hydrophobic residues (H-region), and a region of mostly small uncharged residues containing a characteristic cleavage site recognized by a specific signal peptidase (SPase) (von Heijne, 1984; von Heijne, 1989; von Heijne, 1990). It is this characteristic site that holds the key in cleavage of a secretory protein by either of the two SPases, Type I or Type II. Various types of signal peptides are reported in bacterial systems among which secretory signal peptides (Sec type), Twin arginine signal peptides (TAT type), lipoprotein signal peptides (Lipo type), pseudopilin-like signal peptides, and bactericin and pheromone type signal peptides are most prevalent (Tjalsma *et al*., 2004). However, mainly the first three types of signal peptides (i.e., Sec type, TAT type, and Lipo type) are common in gram-positive bacteria (Pallen *et al*., 2003). Gram-negative bacteria execute protein secretion employing specialized secretion machineries such as type I (ABC transporters), type III (flagellar-type), and type IV (conjugation related)

secretion systems (Papanikou *et al.*, 2007; Saier, 2006). Sec type and Tat type signal peptides are cleaved by Type I SPase, whereas Lipo type signal peptides are cleaved by Type II SPase (Storf *et al.*, 2010).

Tremendous advancement in genome sequencing technology has yielded complete genome sequences of a broad range of bacterial population. Automated prediction of secretomes has generated a lot of interest. Prediction of the signal peptide-containing genes, along with their cleavage sites in the completely sequenced bacterial genomes, have been achieved by employing various algorithms such as Hidden Markov Model (HMM), Neural Network (NN) (Bendtsen *et al.*, 2004), and Support Vector Machines (SVM) (Vert, 2002). There have been various web-based servers that employ these algorithms and use specialized programs to predict the secretomes accurately in a given genome. Some of the frequently accessed programs include Signal P, Signal-CF, SIGCLEAVE, Predisi, SPEPLip, SecretomeP and Phobius.

Mammoth genomic data produced from HMP and MetaHIT has provided immense scope to unravel the complex interactive strategies employed by human gut-associated microbes. Study of codon and amino usage patterns, expression behavior, and functional classification of the predicted secretomes among the bacterial masses might confer fruitful information pertaining to their acclimatization in human gut.

Molecular evolution of secretory proteins of human gut-associated bacteria is another important aspect that demands to be investigated to resolve the mysteries of bacterial adaptation and co-evolution in human gut. A reliable index of genetic drift over evolutionary time is the ratio of Ka (non-synonymous substitutions per site) to Ks (synonymous substitutions per site) for a large set of orthologous genes, based on comparisons of related species. The terms Ka/Ks and dN/dS are often used interchangeably. But for the computational purpose, the ratio of non-synonymous (Ka) to synonymous (Ks) nucleotide substitution rates is frequently used as an indicator of selective pressures on protein-coding genes. Ka/Ks ratio reflects the rate of adaptive evolution against the background rate (Hurst, 2002). This parameter has been widely studied in the analysis of adaptive molecular evolution, and is regarded as a general

method of measuring the rate of sequence evolution. To study the impact of selective pressures on the pattern of genetic divergence, it is necessary to find out the pairwise ratio of Ka/Ks between the orthologous gene pairs. Different regions of a single gene can be exposed to varying selective pressures (Hurst and Pal, 2001). In these cases, calculating Ka/Ks over the entire length of the gene does not provide a detailed picture of the evolutionary constraints associated with the gene. Hence, a finer analysis of the Ka/Ks ratio by using sliding windows of different sizes provides minute details of the selective constraints acting on specific positions of a gene segment. Proper assessment of evolutionary signatures of secreted protein components in bacterial residents of human gut might provide meaningful know-how pertaining to complex communication tactics and adaptive policies opted by the microbes for successful residence in human intestinal environment.

Human beings are susceptible to various microbial pathogens that severely affect their health. Besides being exposed to external pathogens, humans harbor wide range of bacterial communities (members of human microflora) associated with manifestation of severe diseases (Guinane and Cotter, 2013). Investigations of the human genome have led to tremendous advances in the field of biomedical science and drug discovery. Disease etiologies involve the interaction of the human body, external environment and the pathogenic microbes that infect human beings.

Proper identification of drug targets in any pathogenic organism is the most vital step in drug discovery process. Rapid progress in genome sequencing technologies and advances in bioinformatics and cheminformatics based research domains have provided a massive scope for enhancement of drug discovery technologies. Availability of genomes of both the host and concerned pathogen provides a platform for subtractive genomics based drug target identification in concerned pathogen (Allsop, 1998; Stumm *et al*., 2002). Subtractive genomics based approach involves subtraction of host genome from the pathogen while screening the tentative targets in the pathogen. Such a methodology ensures that the drugs, targeted against the essential gene components present solely in the

pathogen, do not interact with the human genes. Computational tools have made it easier to filter the unique essential genes in concerned pathogens that are associated with the robustness and viability of the infective organism (Damte *et al*., 2013; Amineni *et al*., 2010).

Certain bacterial members residing in human gut have been reported to be associated with severe diseases like colorectal and gastric cancers, inflammatory bowel diseases (IBD) and ulcerative colitis (Guinane and Cotter, 2013; Sekirov *et al*., 2010). *Helicobacter pylori* is an intimidating pathogen that resides in human gastrointestinal tract and causes gastritis, peptic ulcers and gastric adenocarcinoma on acquiring pathogenic phenotype (Graham, 2000; Neelapu and Pavani, 2013). Pathogenic members of *Fusobacterium* have been reported to be associated with diseases like ulcerative colitis, colonic inflammation and colorectal cancer (Rajilic-Stojanovic *et al*., 2013; Castellarin *et al*., 2012). Availability of complete genome sequences of numerous human gut-associated bacteria and enormous genomic data produced from HMP and MetaHIT provide ample opportunities to pursue investigations at the genomic and proteomic levels and screen potential drug targets in pathogenic microbes associated with human gut.

This PhD thesis has been envisaged and developed at the Bioinformatics Facility, Department of Botany, University of North Bengal and Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta. The main research motive of this work has been to investigate the genomic and proteomic traits of human gut-associated microbes and unravel the enigma of complex interactions that the bacterial members exhibit with human host.

Objectives of this thesis work:

Explore the mode of codon usage: Comprehensive investigation of codon usage signatures of various bacterial genera residing in human gut would be carried out extensively. Our approach would aid in proper elucidation of the major forces that influence the codon and amino acid usage patterns in the gut microbes. We also intend to compare the codon usage patterns of the concerned bacterial members with human host which might provide a vivid portrait about the adaptive strategies employed by these organisms for successful residence in human gut.

Amino acid usage investigations: We plan to explore the complex amino acid usage behavior of the bacterial members that dwell in human intestinal niche and identify the probable determinants that govern the observed patterns. Correspondence analysis of amino acid usage data would also be executed to probe the apparent causes behind the functional adaptations of encoded proteins.

Comparative genomic analysis: Comparative genomic analysis of various crucial bacterial genera of the human gut would be performed. Specific and characteristic genomic and proteomic signatures of several concerned bacterial genera would also be plowed into. The pan– and core genomic traits of every concerned genus would be scrutinized with a motive to elucidate the adaptive strategies of the microbes within human intestinal niche. Proper know-how of the core genomic architecture of various bacterial members of human gut would provide resource for a better understanding of the 'acclimatization policy' adopted by these groups of microorganisms.

Profiling of Carbohydrate-Active enZymes (CAZymes): We also plan to execute extensive profiling of the special sets of enzymes called Carbohydrate-Active enZymes (CAZymes) that are prevalent in the gut microbes and are associated with the breakdown of complex carbohydrate moieties and glycoconjugates (Cantarel *et al*., 2009). CAZymes break down undigested carbohydrate components of the human diet and facilitate proper digestion of the human host. The carbohydrate components also supply nutritional sources for the gut microflora to thrive and vibe in the human intestinal niche. Thus, proper analysis of the CAZymes in the intestinal microbial pool might render crucial information regarding the metabolic apparatus of these organisms and their close concord with human host.

Comparative secretome analysis: *In silico* prediction and characterization of secretomes, the complete set of secreted proteins, in various bacterial genera of the human gut would be executed. Secretomes are believed to execute several crucial 'remote-control' functions associated with cellular communication and cross-talks (Tjalsma *et al*., 2004; Ranganathan and Garg, 2009). Thus, proper identification and know-how of the secretome sets in various microbes of

human gut might provide a better elucidation of the host-microbe interactions and adaptive tactics employed by these groups of microbes. Amino acid usage tendencies of the secretomes would also be investigated to explore any bias in usage of amino acids among these special sets of proteins. Functional characterization of the secretomes, based on analysis of Clusters of Orthologous Groups (COG) category, would be accomplished to revisit the functional implication of the secretomes. We also intend to compare and analyze the amino acid biosynthetic cost of the secretomes, with respect to the non-secretomes.

Assessment of evolutionary selection on secretomes: Evolutionary signatures of the secretomes, with respect to the non-secretome sets, would be investigated with an intention to assess the forces of selection that might be operative on the secretory machinery of the bacterial systems.

Identification of potential drug targets in pathogenic members of the human intestinal microflora: Microflora of the human intestine has been reported to be associated with several pathogenic manifestations (Canny and McCormick, 2008). *Helicobacter pylori* is an alarming pathogen that resides in human gastrointestinal tract and has been reported to associated with severe intestinal disorders that include gastritis, peptic ulcers and gastric adenocarcinoma (Graham, 2000; Neelapu and Pavani, 2013). Identification of potential drug candidates has been achieved in various strains of *H. pylori* (Sarkar *et al*., 2012; Neelapu and Pavani, 2013; Nammi *et al*., 2016). However, *H. pylori* 35A strain, a pathogenic resident of human gut, still demands to be explored and characterized from the drug discovery aspect. We aim an extensive genomic/proteomic characterization of the *H. pylori* 35A strain for apt identification of putative therapeutic drug targets, which might provide a pedestal for successful drug development against *Helicobacter pylori* associated infections.