Metagenomics of the Human Intestinal Tract

Humans live in constant association with microbes. The number of microbes that are present on surfaces and in cavities of our body largely exceeds that of our own cells and the number of genes they encode largely exceeds that of our own genes. This complex and dynamic microbiota has a profound influence on human physiology and nutrition. Defining this dynamic diversity represents the next frontier of genomics. To progress towards this ambitious goal we focus on the microbiota of the intestinal tract, which is the most complex and plays a particularly important role in human health and well-being. The ensemble of the genomes of human-associated microorganisms represents the human metagenome. A detailed understanding of human biology will require not only knowledge of the human genome but also of the human microbial metagenome.

Our intestinal microbes

The microbes of the human intestine can reach up to ten trillion cells and represent a weight of two kilograms, exceeding that of our brain. They help us digest food, they synthesize vitamins and amino acids that are needed by our body, they protect us by educating the immune system to distinguish friends from foes. Many different diseases originate from microbial disorders. This is naturally the case of infectious diseases affecting the digestive system. But chronic diseases, which are steadily increasing in the modern societies, have also been associated with unusual changes in microbiota. Even the apparently psychological disorders such as autism may be related to an overgrowth of certain bacteria living in the intestine.

The project objectives: association of bacterial genes with human health and disease

The central objective of our project is to establish associations between the genes of the human intestinal microbiota and our health and disease. We focus on two disorders of increasing importance in Europe, Inflammatory Bowel Disease (IBD) and obesity. The incidence of IBD has been increasing constantly during the past decades in Western Europe, and this dramatic trend is now observed in Eastern Europe as well. The global epidemic of obesity is well recognized and imposes a huge and rapidly growing challenge for the public health services.

To reach our central objective we carry out and integrate a number of different activities. First, we established an extensive reference catalog of microbial genes present in the human intestine. Second, we developed bioinformatics tools to store, organize and interpret this information. Third, we developed tools to determine which genes of the reference catalog are present in different individuals and at what frequency. Fourth, we gathered cohorts of individuals, some sick and some healthy, and determined for most which genes they carry. Fifth, we developed methods to study the function of bacterial genes associated with disease aiming to understand the underlying mechanisms and host/microbe interactions.

Our project must be integrated in the world we live in. For this purpose, we actively participate in the International Human Microbiome Consortium (IHMC), carry out transfer of technology to industry and help present the information about the project to the general public.

Catalog of genes of intestinal microbes – our other genome

We established a broad catalogue of microbial genes from the intestinal tract. This was achieved by the cutting edge sequencing technology, which allows to generate tens and even hundreds of millions short sequences in parallel for any DNA sample. We determined a total of some 540 Gb of DNA sequence prepared from stool samples, a value approaching that of 200 human genomes. In a break-through manner, we succeeded to connect these short snippets into much longer DNA stretches, where we could identify the genes present in the intestinal microbes.

We analyzed samples from 124 individuals that participate in our studies. They were of Danish and Spanish origin, some were healthy and some sick, suffering from IBD or obesity. In this way, we expected to identify the largest possible gene number, not missing those that could possibly be less frequent or even absent in a given group of individuals. An extensive bio-informatics analysis has shown that there is a staggering number of some 3.3 million different genes among the individuals that we analyzed, 150-fold more than in our own genome! We have identified at least 85 % of all the frequent genes that the 124 individuals carry, the value determined by an appropriate statistical analysis. Some 99 % of the genes were of bacterial origin, in keeping with the predominance of bacteria among the intestinal microbes. From the gene number we deduce that there are at least a 1000 bacterial species frequent in our gut.

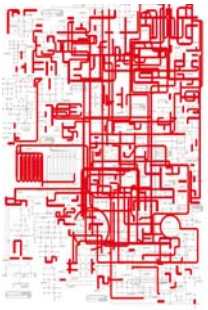
How many of the 1000 are present in each individual? We find that a person carries, on average, 540 000 genes, a value that corresponds to some 160 species. Inevitably, different individuals have many of the bacterial species in common - there are no more than a 1000 to go around and everyone has at least 160. We found some 60 species that are present in at least half of the individuals of the cohort, in keeping with the fact that about 40 % of the genes of each individual are present in at least 50% of other individuals.

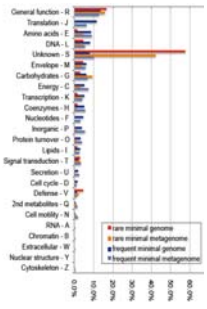
Does our catalog contain the genes from individuals other than those that we have studied? The answer is definitely yes - we find over 80 % of the sequences from 18 US individuals and over 70 % from 12 Japanese individuals, determined in previous studies of a smaller scope in the catalog. Clearly, a majority of genes of the human intestinal bacteria are well represented. Collectively, they have been dubbed “our other genome” on the cover of a recent Nature issue where these results were published.

Information organization and analysis

Sequence information must be interpreted (annotated) in terms of genes, proteins and the functions they perform. An automated sequence annotation pipeline was developed and used to analyze sequence data from the human intestinal metagenome.

We identified over 19000 different functions in the gene catalog that we established. The statistical analysis indicates that we have captured essentially all of the functions present in our 124 samples and thus have an exhaustive view of the genetic potential of the bacteria from the human gut. A large proportion of the functions, over 5000, were never found before. This illustrates the novelty that our analysis has revealed.



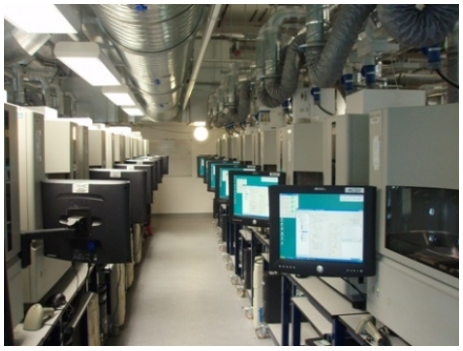


We also identified some 6000 functions that are present in every individual of our cohort. We suggest that they constitute “minimal metagenome”, which may be required for the proper function of the human gut microbiota. Among these are, as expected, the

functions that our genome lacks, such as the capacity to degrade the fibers present in our food and thus extract more energy from it or to synthesize vitamins and amino-acids essential for us. Interestingly, we have very little information for many of the minimal metagenome functions and our findings should prompt their study.

Beyond the minimal metagenome, we defined the set of 1200 functions as required for any bacterium to strive in the human gut and suggest that they represent the "minimal gut genome". About a half are present in most bacteria with sequenced genomes and are necessary for the bacterial life. A large number, however, are found only rarely among the bacteria with sequenced genomes and may well be specific for the gut bacteria. Their study should lead to a much better understanding of our microbial companions than we presently have.

Microbial genes in different individuals



The next challenge we addressed was to determine which genes of the catalog are present in each individual and with what frequency. This enables searching for association between the genes and the disease. For this purpose we followed a 2-pronged approach, using either the very high throughput DNA sequencing or the DNA arrays.

Very high throughput sequencing generates short sequence "tags", that originate from genes present in the sample and thus in our gene catalog. We developed procedures to map the tags onto the catalogue genes, and therefore to count the genes present in each individual. Routinely, we generate some 30 millions tags for each sample and can thus count genes with a high accuracy, as their number is substantially lower (slightly above half a million, on average). We developed the bio-informatics procedures required to determine efficiently the gene frequency in our samples. Almost 400 individuals were analyzed in this way.

We also developed arrays that allow to measure gene frequency and, importantly, also gene expression. One is carried out by analyzing the DNA and the other the RNA from each sample. The present-day technology cannot accommodate all of the catalog genes on a single array (two are required), which renders the routine use of the arrays that contain the complete gene set laborious and costly. We thus also developed a simplified version, that contains the most informative genes on a single array and developed the protocols for its efficient use. This array has already been used for the analysis of over 150 individuals.

Patient cohorts and microbial profiling

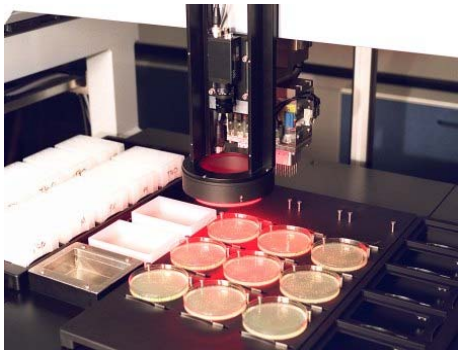
IBD includes two different pathologies, Ulcerative colitis (UC) and Crohn's disease (CD). We planned to compare patients in remission with a similar number of healthy individuals, generally from the same family (30 for each disease). The clinical part of the UC study has now been completed. The individuals have been enrolled into the study, they were clinically examined and their samples (blood and stools) obtained and analyzed. We anticipate that the CD study will reach the same stage before the end of the year, as almost 90% of the individuals have been enrolled.

The bio-informatics comparison of the bacterial genes present in the UC patients and healthy controls has begun and the first results indicate that some bacterial species differ in the two groups. If fully confirmed, this may well lead to a break-through in our understanding of this disease.

Obesity is associated with a number of co-morbidities, such as diabetes, hypertension and cardiovascular diseases. These co-morbidities are found typically in individuals with visceral rather than general obesity. We planned to compare 60 individuals of each type with 60 lean individuals. As in the UC study, the clinical part has been completed. The individuals from a large cohort, Inter99, established in 1999 in Denmark, were enrolled in our study. They were examined at the entry in the study and the observation collected can be compared with those obtained 5 and 9 years ago for all of them. Their stool and blood samples were collected and examined. Once again, the first bio-informatics analyses point to the differences in bacterial species between the obese and lean individuals, that could lead to a breakthrough in the comprehension of this ever-increasing disease.

Bacteria-host interactions

We focused on the genes that may be involved in the interaction between bacteria and ourselves. To identify these genes we developed procedures allowing to monitor the response of human cell lines when brought in contact with the genes present in gut bacteria. A two-pronged approach was followed. On the one hand, we constructed a large collection of genes from intestinal bacteria in a standard E. coli cloning host. The collection comprises over 200 000 clones, and a total of over 8 million genes from intestinal bacteria. On the other, we established 16 different screens, based on the human cell lines carrying various reporter genes. We validated our approach in a pilot study with 5000 bacterial clones, where a high-throughput robot was used to carry out 25 000 individual tests. About a dozen clones were found to induce a significant cellular response.



We further tested, in culture, the effect of 5 clones on the dendritic cells (DC), either directly or indirectly via intestinal epithelial cells (IEC). DCs are thought to be among central elements of the immune system response. The metagenomic clones differently regulated gene expression in IECs, and their impact on IECs clearly conditioned DCs response. In addition, one clone directly affected the DCs. Analysis of these clones may lead to a novel understanding of interactions between the bacteria that we host in our gut and us.

Technology transfer

Some of the clinical studies of IBD involve MetaHIT industrial partners. We centered on the one that concerns the effect of probiotics on the stability of microbiota in the UC patients, thought to be lower than in the healthy individuals. The study involved 50 patients and 10 healthy controls. One group of patients consumed daily a fermented milk product containing probiotic strains for three months, while another group took a product without the strains for the same period of time. Stool samples were collected at the onset and the end of the study and analyzed by the high throughput DNA sequencing. A preliminary analysis of the fluctuation of the bacterial gene abundance showed the expected high stability among the healthy individuals and a significantly lower stability among the patients taking placebo. Very interestingly, the patients that consumed the probiotics had an improved stability, similar to that of the healthy individuals. A number of follow-up studies are planned at present, in close interaction with the MetaHIT project.

Communication and Coordination in the Human Metagenome field

To promote the necessary international cooperation and coordination MetaHIT took an active role in the establishment and functioning of the International Human Microbiome Consortium (IHMC). The MetaHIT coordinator served as a co-chair of the IHMC and several partners participated in the IHMC working groups, related to genome sequencing of the gut bacterial strains, and the controlled release of the clinical data.

MetaHIT organized the First International Conference on Human Metagenomics. The Conference took place on March 1-3 in Shen Zhen, China, the home town of the MetaHIT partner BGI. It gathered over 200 participants from 23 countries, representing large Human Microbiome-related projects over the four continents.

Information of the general public about the field, the project and its achievements, is high on the MetaHIT list of priorities. One of the privileged ways is *via* its Web site (http://www.metahit.eu). MetaHIT is now also present on 2.0 platforms such as Facebook (MetaHIT page) and Twitter (www.twitter.com/metahit), thus enabling real-time updates accessible through popular media. We also opened a dashboard on Netvibes (www.netvibes.com/metahit) gathering Human Microbiome live news. This is the first known effort in centralization of information in that specific field on the internet.

Communication with newspapers, magazines, radio, and television has been pursued attentively. MetaHIT attracted attention of the major media at the release of the Nature cover publication in March, a collection of articles is accessible on our website (http://www.metahit.eu/index.php?id=205).

The MetaHIT project was chosen to be part of the European COMED project (http://www.comed-project.org/health-research-projects/meta-hit.html). Several partners took part in the production of a 8-minute movie, highlighting our research by focusing on the Crohn's Disease studies.

Expected achievements and their impact

The expected final achievements are the establishment of the methodology to characterize individual intestinal metagenomes and the discovery of associations between bacterial genes and human disease. This should pave the way for the development of novel diagnostic and prognostic tools, based on microbial genes, aiding preventive and personalized medicine. Furthermore, the detailed description of the human intestinal microbiota, its dynamics and its interaction with the human host will lead to a much more complete understanding of human biology.

Last but not least, by combining methodological and conceptual advances our project will result in, we expect to open avenues for reasoned modulation of our microbiota. The ability to characterize individual intestinal microbiota and follow its evolution with time, coupled to the understanding of the microbe/host interactions, should allow to explore the effects of factors such as food, environment and age on the dynamics of our microbial populations, and to develop interventions to optimize these populations. This should open novel possibilities to improve human health and well-being. In a nutshell, we expect MetaHIT to open new and revolutionary avenues in medicine, that will target our other genome.