Genomic reconstruction of metabolic pathways for the utilization of carbohydrates in the human intestinal microbiota

Arthur A. Murtazin[1], Semen A. Leyn[1,] Dmitry A. Rodionov[1]

[1]*Institute for Information Transmission Problems, Moscow, Russian Federation*

e-mail: aa.murtazin@gmail.com, semen.leyn@gmail.com

rodionov@iitp.ru

Background: Human microbiome is a set of all symbiotic, commensal and parasitic microorganisms that inhabit the human body. The total number of bacterial cells in the human body can vary from 10 to 100 trillion pcs, and the majority of microbiota dwells in the gastrointestinal tract.

Human intestinal microbiota plays a key role in utilization of carbohydrates. The human genome contains less than 20 genes encoding enzymes for cleaving complex polysaccharides. At the same time, the genome of only one representative of the intestinal microbiota, *Bacteroides thetaiotaomicron*, contains genes encoding more than 250 enzymes involved in the metabolism of complex polysaccharides. It has been shown that the members of gastrointestinal microbiota are capable of cleaving a significant amount of complex polysaccharides of plant origin, which are ingested with food. Among these polysaccharides are fructans, xylans, pectins and arabinans. In addition, gut microbiota is capable of cleaving a number of polysaccharides that are synthesized by cells of the human body, thus making them easier to recycle. The ability of some inhabitants of the gastrointestinal tract to cleave chondroitin sulfate, heparin and hyaluronic acid has been demonstrated. Also, representatives of gut microbiota can metabolize simple sugars and disaccharides such as fructose, rhamnose, N-acetylglucosamine and human milk oligosaccharides that are coming with food.

Results: We use a comparative genomics-based approach to assess *in silico* the metabolic potential for utilizing key carbohydrates of nearly 500 members of the human gut microbiome. For functional gene annotation and pathway reconstruction, we used three comparative genomic techniques: (i) homology-based methods (BLAST); (ii) genome context analysis (PubSEED); (iii) co-regulation by the same transcription factor (RegPrecise). Upstream glycohydrolytic machinery was mapped using CAZy. We report detailed

reconstructions of metabolic pathways and regulons for utilization of more than 20 carbohydrates and polysaccharides. We investigated certain pathway variations that are quite common in metabolic pathways of sugar utilization. Analyzing presence or absence of key transporters and glycosyl hydrolases we have been able to distinct ability to utilize certain oligosaccharides (and polysaccharides) from the ability to only utilize corresponding monosaccharide. The absence of monosaccharide transporters in many cases could be explained by the presence of other catabolic/utilization pathways that can produce this monosaccharide intracellularly. For example, galactose could be supplied from beta- and alpha-galactosides, raffinose, melibiose, and lactose, while glucuronate is released from the breakdown of glucuronides. We identified sets of carbohydrate active glycosyl hydrolases that are involved in breakdown of oligosaccharides (and polysaccharides) into monosaccharides. Numbers of predicted glycosyl hydrolases range between 20 and 66 enzymes per genome. We further reconstructed catabolic pathways for utilization of various oligosaccharides including lactose, maltose/maltodextrin, sucrose, raffinose, melibiose, chitobiose, lacto-N-biose, alpha-galactosides, beta-galactosides, beta-glucosides, fructooligosaccharides, rhamnogalacturonides, rhamnosides, glucuronides, xylosides, arabinosides, fucosides, hyaluronate and sialic acids. We also report non-orthologic displacements of enzymes in certain pathways that reflect evolutionary diversity in these metabolic pathways.

Conclusions: We performed genomic reconstruction of sugar utilization metabolic pathways in over 500 members of human gut microbiota. The reconstructed regulatory networks can provide new insights on metabolic capabilities of these bacteria and be used for spreading results to even larger number of species and strains. We later plan to test the prospects of using these results in clinical practice.

Acknowledgments: This research was supported by the Russian Science Foundation (grant #14-14-00289).