

Examining composition of Russian human gut microbiota by assessing relative abundance of functional and taxonomical units

Alexander V. Tyakht¹, Anna S. Popenko¹, Maxim S. Belenikin^{1,3}, Ilya A. Altukhov^{1,2}, Dmitry G. Alexeev^{1,2}

¹Research Institute of Physico-Chemical Medicine (RIPCM), Malaya Pirogovskaya 1a, 119435 Moscow, Russia

²Moscow Institute of Physics and Technology (MIPT), Institutskii per. 9, 141700 Dolgoprudny, Moscow Region, Russia

³Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilov st. 32, 119991 Moscow, Russia

Introduction

Russian Metagenomic Project is an initiative lead by a consortium of Russian scientific and medical institutes targeting study of human gut microbiome of Russian population. During Phase I of project, fecal metagenomic samples of 132 subjects were sequenced on ABI SOLiD sequencer and analyzed bioinformatically. The composition was derived by assessment of taxonomic and functional relative abundance, allowing linking with various medical and ethno-geographical factors. Taxonomic profiling revealed clustering of types of microbiota into clusters similar to previously described enterotypes, although the third (mixed-type) enterotype seems to demonstrate more complex sub-structure. The differences were simultaneously examined using abundance of prevalent microbial genes grouped into functional categories. Within the frame of analysis, Russian samples were compared with metagenomic data sets from MetaHIT and HMP projects. For selected samples, taxonomic results of whole-genome approach was compared to 16S rRNA sequencing on 454 sequencer.



Fig. 1. Geography of samples

Methods

Analytical pipeline

Using alignment software (Bowtie), metagenomic reads were mapped to two reference sequences sets: MetaHIT gene catalog of 3 mln prevalent human gut microbial genes [1] and set of known genomes (collected from HMP and NCBI). Normalized coverage depth and width (calculated with BEDtools) were used to estimate relative abundance of taxonomic and functional units. Statistical analysis and visualization were performed using R: principal components (PCA) and between-class analysis (BCA), k-medoids and hierarchical clustering.

Taxonomic profiling: prevalent genera

Quantitative bacterial composition was computed at genus level by summing coverage of genomes within each genus. This level of taxonomic resolution corresponded to short read length. Genomes with percent of covered length below threshold value were discarded. The resulting taxonomic signature was used as feature vector for statistical analysis.

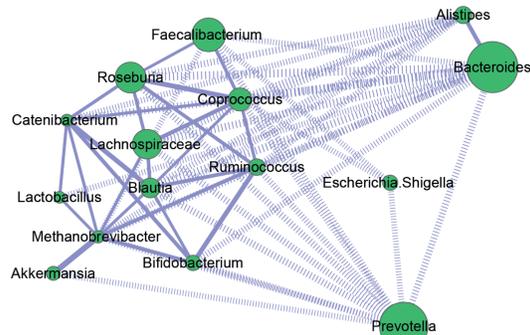


Fig. 2. Correlation graph of top bacterial genera abundance: cooperation and opposition (Cytoscape).

Functional profiling: enriched COGs

Genes from gene catalog were binned according to their functional annotation (Clusters of Orthologous Groups). In their turn, COGs were grouped by activities of particular interest: genes associated with antibiotic resistance, transport, vitamin metabolism and others. Enrichment of certain COG points to high potential of corresponding metabolic link in total metabolic network of microbiome.

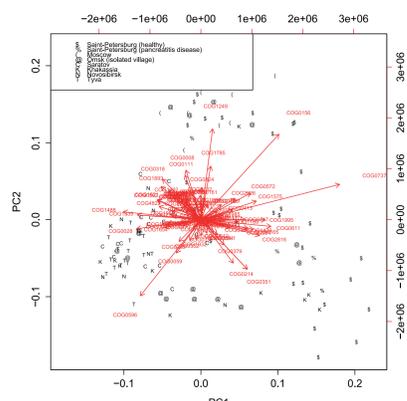


Fig. 3. PCA plot for enrichment of functional groups associated with vitamin metabolism. Symbols correspond to various geographic and medical groups of samples.

Assessing specific metagenomic potential

Additionally, to examine enrichment of interesting potentials in microbiome, we employed specific reference sequence sets for mapping:

- Antibiotic-resistance associated genes: ARDB
- Bacterial toxins: DBETH
- Horizontal gene transfer: sequences from [3]
- Transporters: TCDB, TransportDB

For nucleotide reference sets, Bowtie was used, while for protein sets, short reads were mapped using RAP-search. Selection of specific reference sets can reveal clustering of samples – in case when there is no clustering at global scale of all genes.

Results

Clustering reveals links of microbiota composition with ethnogeographic and medical factors

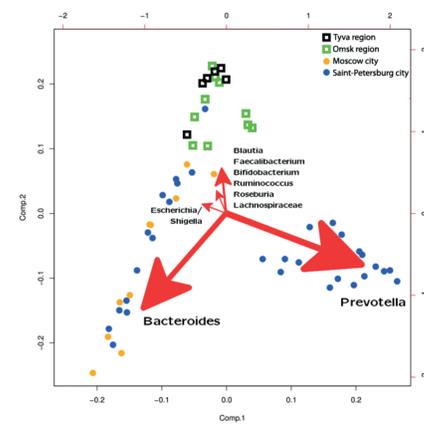
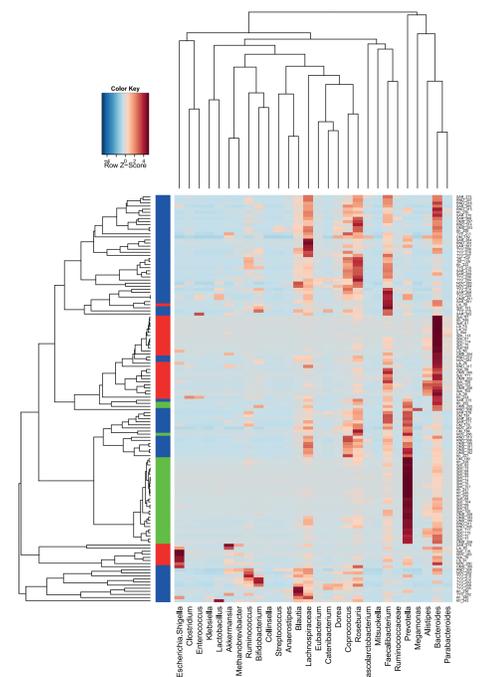


Fig. 5. Bacterial composition of city vs country-side metagenome (PCA plot for genera abundance, samples from two large cities and two remote areas).

Fig. 6. Hierarchical clustering of samples based on genera abundance. Rows are samples, columns – genera. Three-colored column on the left reflects k-medoid clustering. The resulting organization resembles three enterotypes [4], with Bacteroides and Prevotella dominating each its own cluster and the third being more “pluralistic” (many abundant species without clear leader).



Diversity of world gut metagenome

Within the frame of world global metagenomic context, we compared Russian 132 samples together with 124 MetaHIT and 139 HMP short-read metagenomic samples [2] using statistical analysis and visualization tools.

Extension of gene catalog: novel genes

Short reads were assembled *de novo* using SOLiD denovo tools. After extraction of ORFs and translation, the resulting protein sequences were aligned to MetaHIT gene catalog using BLASTP. Novel 33,688 thousand proteins with length > 100 aa were obtained. Removal of redundancy yielded 32,530 proteins that can be considered as 1% novel addition to catalog of gut microbial genes.

Genera abundance: whole-genome vs 16S rRNA

For comparison with other methods of taxonomic profiling, 16s rRNA genes were sequenced on 454 for selected samples. The produced long reads were classified using RDP classifier and compared with our short-reads mapping approach, as well as with abundance values resulting from mapping those short reads to 16S reference databases (SILVA, Greengenes). Across these approaches, several top genera common for all bases remained the same, but quantitative correlation was low and many genera from 16S database didn't have whole-genome representative.

Acknowledgements

We thank Prof. Vadim Govorun who participated in study design and coordination. This work was supported by State Contracts 16.512.11.2111, 16.552.11.7034.

References

1. Qin, J. et al: A human gut microbial gene catalog established by metagenomic sequencing. Nature 2010, 464:59-65.
2. Human Microbiome Project Data Analysis and Coordination Center [http://www.hmpdacc.org]
3. Smillie, C. et al: Ecology drives a global network of gene exchange connecting the human microbiome. Nature 2011, 480: 241-244.
4. Arumugam, M. et al: Enterotypes of the human gut microbiome. Nature 2011, 473:174-180.

For more information on Russian Metagenome Project, check out other posters by our group:

- MALINA - a Web-service for human gut microbiota whole-genome metagenomic reads analysis

- Human gut microbiota analyzed by mass-spectrometry and sequencing

