Combinatorial chemistry discriminating analysis of complex microbial systems with restricted site tags (RST)

Alexey Kutsenko[1], Veronika Zabarovska[1], Lev Petrenko[1], Tore Midtvedt[2], Ingemar Ernberg[1], Eugene R. Zabarovsky[1]

Keywords: biodiversity, gemone, microorganism, human microflora, combinatorial chemistry

1 Introduction.

The current strategies for mapping and sequencing are clearly not meeting the challenge of high-throughput comparative genomics. Alternative and complementary strategies need to be developed, and it is imperative now to find cost-effective and convenient methods that allow comparative genomics projects to be undertaken by a wide range of laboratories.

We develop a new robust and high efficient technique for large scale scanning of genomes in complex multiorganisms mixture on a quantitative and qualitative basis. This comparative genomic technique are currently effectively applying [1] in the area of sequencing of related bacterial strains and species, in order to identify the genomic basis for differences in their biological properties, particularly pathogenicity. Our approach allows analysis of complex microbial mixtures such as in human gut and identification with high accuracy of a particular bacterial strain on a quantitative and qualitative basis.

2 Method and results.

To achieve the aim of large scale scanning of microbial genomes we propose to create restricted site tags (RST) set of a organism: databases containing specie's short sequences surrounding restriction sites - tags, of a particular rare cutting restriction enzymes. We introduced an information value of a rare cutting restrictases (with restriction site of eight base pairs in lengths and more, Tab. 1, Fig. 1) and analyzed RST passports of all selected restrictases. Thus, a comparison of 1 312 tags from available sequenced *E. coli* genomes, generated with the *Not*I, *Pme*I and *Sbf*I restriction enzymes, revealed only 219 tags that were not unique. None of these tags matched human or rodent sequences.

Restriction enzyme	***Pme*I**	***Sbf*I**	*Pac*I	*Fsp*A	***Not*I**	*Sgf*I	*Sgr*AI	*Srf*I	*Sse*2321	*Asc*I	*Fse*I	*Swa*I
Score	**62.5**	**49.5**	37	35.5	**22.5**	22.5	21.5	16.5	16	14	13	-21

Table 1. Information value of recognition sites for rare cutting restriction enzymes in selected 70 microbial genomes.

Thus, RST set for a particular organism represents in fact a unique genomic fingerprint of this specie or strain that is easy to generate. This distinctive feature of the method allows a

[1] Microbiology and Tumor Biology Center, Karolinska Institute, Stockholm, Sweden. E-mail: Alexey.Kutsenko@mtc.ki.se

[2] Department of Cell and Molecular Biology, Karolinska Institute, Stockholm, Sweden.

discriminating different species and even strains from mixture of genomes. From the total DNA fraction of a complex microbial system a total RST set is produced and sequenced. Then an information about the original genomes mixture is extracting by comparison of the total set with the fingerprints of known bacteria (with known genomes) using combinatorial chemistry. At the same time it was shown that short sequences randomly taken from any bacterial genomes fall in clusters when principal component analysis (PCA) is applied [2]. Due to this fact we can suggest the potential of our method to discriminate even between unknown (unsequenced) species.

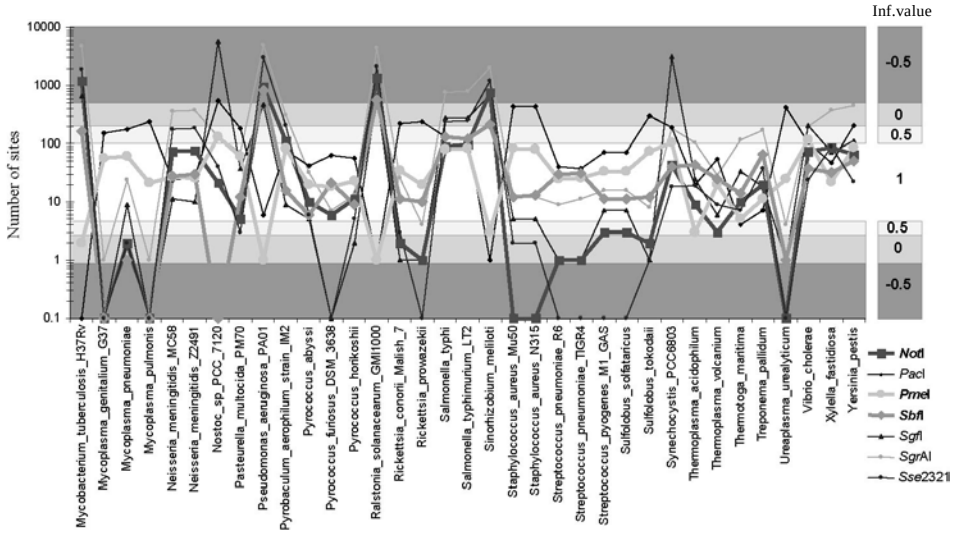

Figure 1. Schematic distribution of recognition sites for rare cutting restriction enzymes in selected completely sequenced bacterial genomes.

The procedure for generating tagged sequences is easily realized experimentally and can be adapted to any rare cutting restriction enzyme. We demonstrated experimentally that the *Not*I tags comprising 19 bp of sequence information could be successfully generated using DNA isolated from intestinal samples. Such *Not*I passports allow the discrimination between closely related bacterial species and even strains. Therefore the approach allows analysis of complex microbial mixtures such as in human gut and identification with high accuracy of a particular bacterial strain on a quantitative and qualitative basis.

The remarkable advantage of our method verified on microbial systems is the ability to identify even strain composition. This gives the opportunity of identifying faint differences between relative organisms, e.g. pathogenic islands.

References.

[1] Zabarovska, V., Kutsenko, A.S. *et al.* 2003 *Not*I passporting to identify species composition of complex microbial systems. *Nucleic Acids Res.*, 31:e5.

[2] Sandberg, R., Winberg, G., Branden, C.-I., Kaske, A., Ernberg, I. and Coster, J. 2001 Capturing whole-genome characteristics in short sequences using a naïve Bayesian classifier. *Genome Res.*, 11:1404-1409.