Comparison of 19 commensal human and animal *Escherichia coli* genomes using oligoarrays

Grasselli E. [1], Benagli C. [1], Gettler B. [2], François P. [2], Boerlin P. [3], Schrenzel J. [2], Piffaretti J.-C. [1]
[1] Istituto Cantonale di Microbiologia, via Mirasole 22A, 6501 Bellinzona, Switzerland
[2] Hôpital Cantonal Genève, Genomic Research Laboratory, rue Micheli-du-Crest 24, 1211 Geneva, Switzerland
[3] University of Guelph, Ontario Veterinary College, Ontario, Canada

Background:
Genome sequencing of many bacterial strains within a species has opened new possibilities in the genomics research field. The comparison of three complete *Escherichia coli* genome sequences, K12, O157:H7 and CFT073, clearly evidenced a common major gene cluster, which represents the *E. coli* core gene pool [Welch *et al.* PNAS 2002]. All remaining genes, which are not shared by these three genomes can be tentatively considered as the flexible gene pool. Most of these last loci are thought to be acquired from an external source, by horizontal gene transfer mechanism.
In order to get a deeper insight into the *E. coli* genome plasticity and to understand the reasons of gene transfers among bacteria, 19 commensal *E. coli* strains collected from diverse hosts (animals and humans) were analysed by an in-house designed oligoarray.

Methods:
61.3% (2'700) of the 4'279 genes of the *E. coli* K12 genome have been spotted on an oligoarray. In order to identify genome differences between the K12 genome and the bacterial strains under investigation, CTP-Cyanine-3 labelled K12 genomic DNA and a CTP-Cyanine-5 labelled test strain DNA were co-hybridised on the oligoarray. The hybridisation signal differences were examined. The resulting log ratios of the fluorescence were normalised and the probability of presence of the signal was estimated for each gene. Divergent genes were meant either as absent or sufficiently different to hinder hybridisation to the related oligonucleotide.

Results:
1'015 genes, 38% of the 2700 genes that were analysed, resulted variable among the 19 commensal strains. Each strain showed a specific variability pattern in the genome, no common pattern was found in bacteria isolated from the same host.
Four-hundred-fifty-eight genes failed to be detected in 25% of the 19 tested *E. coli* strains. These 458 hyper-divergent genes, are clustered in precise regions of the microbial genome, so called hot-spots. A vast majority of these genes (425/458) did not belong to the core genes shared by strains K12, O157:H7, and CFT073. Among the hyper-divergent genes, those involved in defence mechanisms, cell motility and intracellular trafficking and secretion were far more represented than others.

Conclusions:
The genomotyping of 19 commensal *E. coli* strains underlines the partition of the *E. coli* genome in core and divergent genes.
By analyzing hyper-divergent regions, *i.e.* those characterised by higher rates of inter-strain variation, we identified several hot-spots, that do not belong to the core genome. This suggests that some parts of the genome are more frequently rearranged than others.
The bacterial chromosome is a dynamic macromolecule, subject to exogenous and endogenous processes of gene loss, acquisition or mutation. The plasticity of the genome is primarily involved in the maintenance of the genetic diversity and thus is a source of evolutionary adaptation.