# GENOME ANALYSIS: A TOOL FOR THE EXPOLITATION OF INDUSTRIAL BACTERIA

Pier Sandro Cocconcelli, Daniela Bassi, Fabrizio Cappa, Cecilia Fontana, Simona Gazzola, Ester Pietta, Edoardo Puglisi.
Istituto di Microbiologia, Università Cattolica del Sacro Cuore, via Emilia Parmense 84, 29122 Piacenza, Italy
pier.cocconcelli@unicatt.it

*Key words: genome analysis, metabolic pathways, industrial bacteria*

**Introduction.**
Recent improvements in the quality, efficiency, and cost of generation sequencing (NGS) technologies have led to sequence multitudes of bacterial genomes. The interpretation of the resultant sequence information becomes an essential approach to study bacterial biodiversity, to reconstruct the biochemical pathways, to assess the virulence and to analyze the phylogenetic relationship among strains and species.
Aim of this work is to highlight how NGS might provide a powerful tool for deeper understanding of industrially relevant bacteria. Examples will be shown by summarizing data derived from the genome analyses of *Clostridium tyrobutyricum, Lactobacillus rhamnosus, Enterococcus faecium* and *Pseudomonas* spp.

**Methods.** Analysed strains were *Clostridium tyrobutyricum* DSMZ2637 and *UC7086, Lactobacillus rhamnosus* ATCC53103 and ATCC 53103Δ1*, Enterococcus faecium* UC7251, UC8733, UC7267, UC10237, UC8668, UC7256, UC7266, UC7265 and *Pseudomonas spp* UC7153. Genome sequence was made using an Illumina Genome Analyzer Hiseq1000. Quality reader filter and assembly was performed using Velvet 1.2.08. Open reading frames (ORFs) were predicted using Glimmer 3.02. Functional annotations were performed with Manatee and RAST (http://manatee.sourceforge.net - http://rast.nmpdr.org). KEGG (www.genome.jp/kegg) was used to reconstruct biochemical pathways.

**Results.** To show how NGS can provide relevant information on bacterial strains of industrial interest, data deriving from the analysis of 13, genomes will be presented, as summarized in table 1.

*Clostridium tyrobutyricum*, a species involved in cheese spoilage, is of interest for the production of hydrogen from by-products. $H_2$ derives from pyruvate conversion to acetyl-CoA, a reaction catalyzed by pyruvate-ferredoxin oxydoreductase, and metabolic engineering has been prosed as a toll to improve $H_2$ yields. Since the genome of this species was not available, we have sequence the genome of the type strain and of a wild isolate. This resulted in a complete reconstruction of the $H_2$ production pathways and a deeper knowledge on sporulation/germination cycles.

Although *E. faecium* is used in food fermentation and as human and animal probiotic, this species is a recognized human pathogen. In the course of this work we have sequenced the whole genome of six strains from the food chain. Comparison with the genomes of clinical isolates allowed to identify genetic markers able to discriminate safe *E. faecium* from those responsible for human infections.

*L. rhamnosus* ATCC53103, known also as GG, is a strain widely used a human probiotic and, for this purpose, is added to food. Genome analysis of isolated from food has reveled that this strain encountered genome rearrangements, which lead to a 140 kb deletion. This deletion affects the functional properties of this strain as probiotic culture, influencing carbohydrate metabolism, anchoring of surface protein, membrane transport and metabolism of amino acids.

*Pseudomonas spp* UC7153, a psychrotrophic strain isolated from an Italian glacier, is able to degrade polycyclic aromatic hydrocarbon (PAH) at temperature close to 0°C. The genome analysis clarified the its phylogenetic relationship within the *Pseudomonas* genus. Moreover, the functional annotation demonstrated the presence of a variety of genes coding for monooxigenases, potentially involved in the PAH degradation and of genes involved in the adaptation at low temperature. This information opens the way to the industrial use of this strain for bioremediation.

**Table 1.** List of studied bacterial genomes.

| Strain | Species | Genome size (Mb) | Coding genes |
|---|---|---|---|
| DSMZ2637 | *C. tyrobutyricum* | 3.01 | 3002 |
| UC7086 | *C. tyrobutyricum* | 3.07 | 3017 |
| UC7251 | *E. faecium* | 2.71 | 2829 |
| UC8733 | *E. faecium* | 2.73 | 2829 |
| UC7267 | *E. faecium* | 2.74 | 2941 |
| UC10237 | *E. faecium* | 2.60 | 2564 |
| UC8668 | *E. faecium* | 2.91 | 2988 |
| UC7256 | *E. faecium* | 2.83 | 2942 |
| UC7266 | *E. faecium* | 2.92 | 3308 |
| UC7265 | *E. faecium* | 2.88 | 3300 |
| ATCC 53103 | *L. rhamonus* | 2.94 | 2852 |
| ATCC 53103Δ1 | *L. rhamonus* | 2.80 | 2762 |
| UC7153 | *Pseudomonas* spp. | 6.71 | 6070 |

**Conclusions.** Genomic analyses are a potent tool to improve the knowledge of bacteria, and to identify and characterize strains with peculiar features of food and industrial interest.

**Acknowledgements**. This work was supported by the national projects [Genobact (Regione Lombardia), Microbioma (Regione Lombardia), FILIGRANA (MIPAAF)] and by the FP7 project µ-Andes [Ares(2013)364278].