

Development of tools based on UCSC and KEGG for the annotation of the *Photobacterium profundum* genome

N. Vitulo^{*}, A.Cestaro^{*}, A.Vezzi^{*}, S.Campanaro^{*}, F.Simonato^{*}, F.Lauro[§], G.Malacrida^{*},
B.Simionati^{*}, N.Cannata^{*}, D.Bartlett[§] and G.Valle^{*}.

^{*} CRIBI Biotechnology Centre, University of Padova, via Ugo Bassi 58/B, 35131 Padova, Italy.

[§] Scripps Institution of Oceanography, UCSD 9500 Gilman Drive, La Jolla CA, 92093-0202

Keywords. Genomics, Annotation, Genome browser, Kegg, UCSC

One of the critical steps in a genome sequencing project is the efficient data storage and retrieval of the large amount of information produced, which represents the starting point for data analysis and interpretation. We have recently completed the genome sequence of *Photobacterium profundum* strain SS9 and the data have been implemented in a genome browser under the UCSC environment. The UCSC genome browser has been developed at the University of California, Santa Cruz and CRIBI hosts one of their official mirror sites at <http://genome.cribi.unipd.it>. The sequence and annotation information is stored in a MySQL relational database and a web-based tool performs fast visualization and querying of the data. The records are displayed as a series of tracks aligned with the genomic sequence. The *Photobacterium profundum* genome browser contains the ORF prediction obtained by two different programs (Orpheus and Glimmer) and the related non-redundant ORF consensus, the ribosome, tRNA, operons, the clones spotted on the microarray chips, the differentially expressed clones derived from microarray experiments, the orthologous genes on other bacteria, the phage and a prediction of the repeated element on the genome. Figure 1 shows a view given by the genome browser.

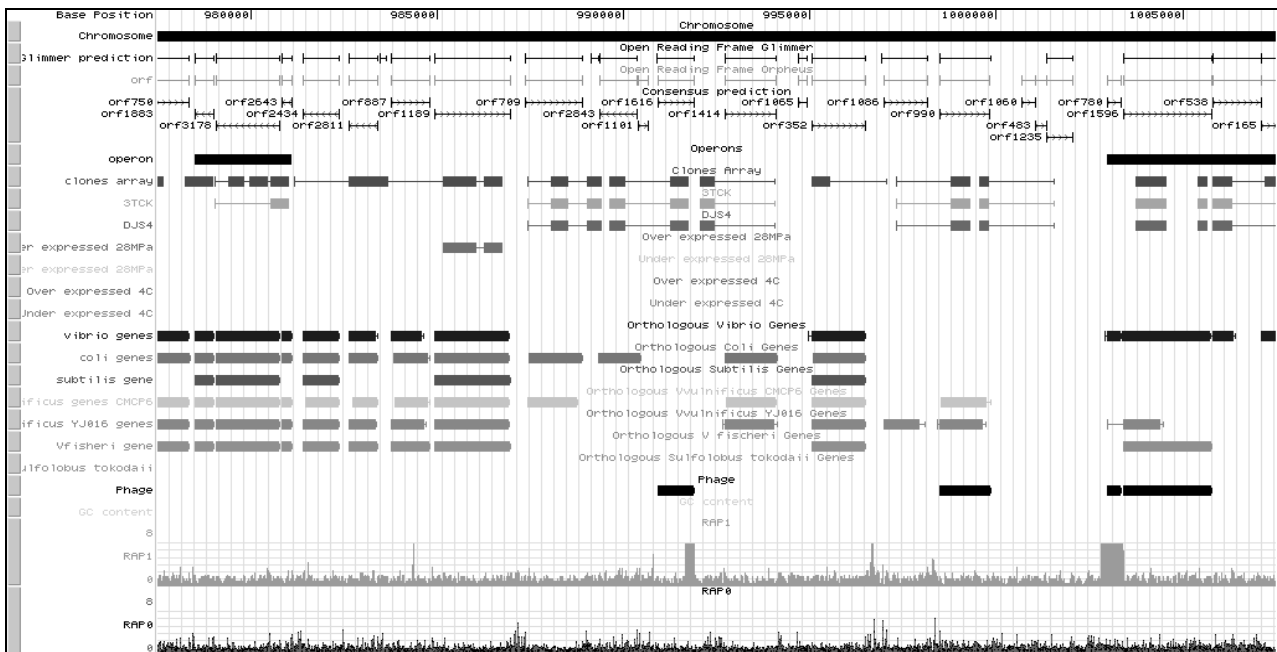


Figure 1. View of the *P. profundum* genome with the UCSC genome browser

The UCSC genome browser has been also implemented with new tools, developed by our group, to manage data that are not available in the original UCSC MySQL structure. In particular, the Kegg information about metabolic pathways and structural proteins has been fully integrated with the UCSC database allowing to visualize on-the-fly the specific maps and the related genes according to their expression pattern. Figure 2 shows the glycolyse map and the flagellar assembly map.

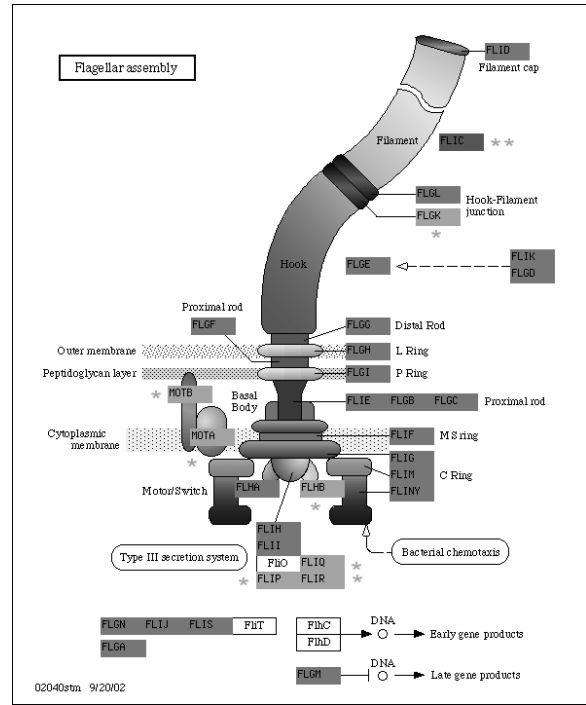
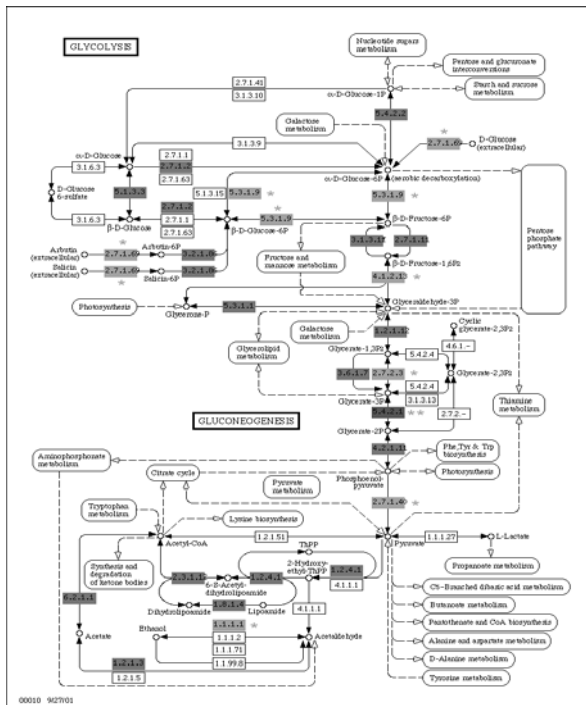


Figure 2. Left panel shows the glycolysis map, the right panel shows the flagellar assembly map. The genes indicated by colored boxes are present in the *P. profundum* genome. Those marked with a ** are the ORF selected from the genome browser, the genes marked with a * are differentially expressed ORF.

Moreover it has been developed a bacterial genome viewer web tool for the visualization of the chromosomes in a ring-like way, which uses the data stored in the UCSC database. Similar utilities have been developed also by other groups, however, we thought that it was useful to produce a tool fully integrated with the UCSC database.

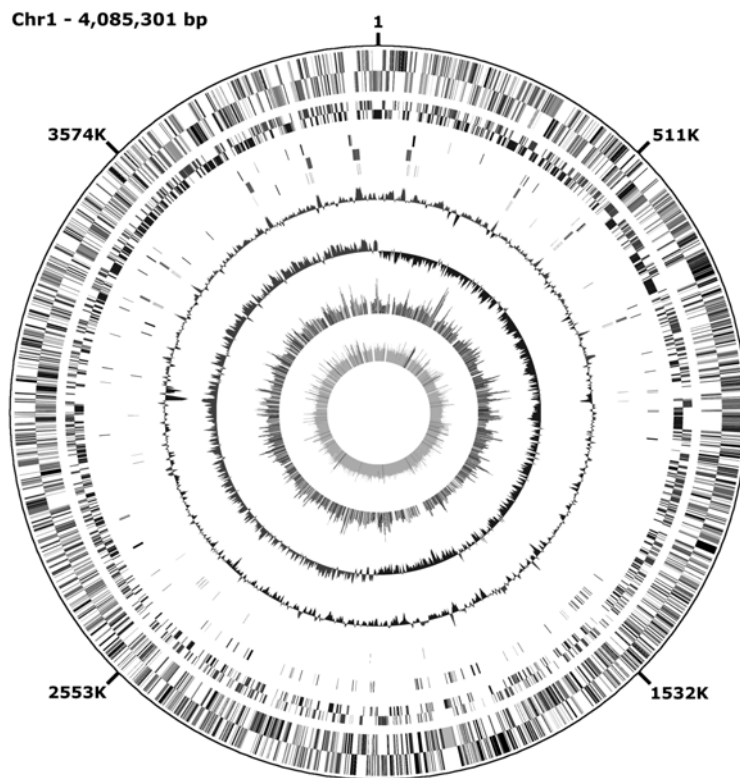


Figure 3. Genomic organization of *P. profundum* strain SS9. From the outside inwards the circle shows: 1-2) predicted ORF on the plus and minus strands; 3-4) *V. vulnificus* YJ016 orthologous genes in both strand; phage and transposon; 6) rRNA operons; 7) tRNA; 8) GC content; 9) GC skew; 10) mean fluorescence value of the microarray clones at 28Mpa; 11) codon adaptation index