Lait 84 (2004) 15–24
© INRA, EDP Sciences, 2004
DOI: 10.1051/lait:2003041

Original article

Whole genome sequencing project of a dairy *Propionibacterium freudenreichii* subsp. *shermanii* genome: progress and first bioinformatic analysis

Guillaume MEURICE[a†], Daniel JACOB[a**], Catherine DEBORDE[a], Stéphane CHAILLOU[b], Annette ROUAULT[a], Pauline LEVERRIER[a], Gwénaël JAN[a], Anne THIERRY[a], Marie-Bernadette MAILLARD[a], Paulette AMET[a], Marc LALANDE[c], Monique ZAGOREC[b], Patrick BOYAVAL[a***], Diliana DIMOVA[a†*]

[a] Laboratoire de Recherche de Technologie Laitière, INRA, 65 rue de Saint-Brieuc, 35042 Rennes Cedex, France
[b] Unité Flore Lactique et Environnement Carné, INRA-CRJ, Domaine de Vilvert, 78350 Jouy-en-Josas, France
[c] Laboratoire de Génie des Procédés et Technologie Alimentaires, INRA, 369 rue Jules Guesde, BP 39, 59651 Villeneuve-d'Ascq Cedex, France

Abstract – Dairy propionibacteria, and especially *Propionibacterium freudenreichii* subsp. *shermanii*, are important in the food industry and biotechnology. Only a few investigations have focused on the complex physiology of this remarkable bacterium, while the physiology of dairy lactic acid bacteria has been extensively studied over the past decades. Here we report the progress of our whole genome sequencing project: 93% of the *P. shermanii* genome was assembled in 426 contigs with low overall redundancy. Our annotation strategy overlaps the sequence finishing step, thus improving it. Annotation of the incomplete genome is performed using ContigBrowser, a bioinformatic tool allowing data management, developed in our laboratory. This resulted in the detection of 2611 putative proteins (data May 2003). Our tool allows an expert annotation by manual verification and curation of functional protein categories after automatic assignment. Our genomic sequence analysis, combined with the already developed physiological, proteomic and metabolomic approaches, will allow researchers to explore the significant potentialities of dairy propionibacteria by providing a comprehensive view of the enzymes and metabolic pathways. This knowledge will allow researchers to explore more effective strategies to enhance the utility of this organism in manufacturing procedures or current industrial processes.

Propionibacterium **/ whole genome sequencing / genomics / genome annotation / bioinformatics**

* Corresponding author: dimova@labtechno.roazhon.inra.fr
† These authors contributed equally to this work.
** Present address: Centre de bioinformatique de Bordeaux (CBIB), 146 rue Léo Saignat, 33076 Bordeaux Cedex, France
*** Present address: Rhodia Food, R & D, BP 10, 86220 Dangé-Saint-Romain, France

Résumé – Projet de séquençage du génome de *Propionibacterium freudenreichii* subsp. *shermanii* : état d'avancement et premières analyses bioinformatiques. Les bactéries propioniques laitières, et plus particulièrement *Propionibacterium freudenreichii* subsp. *shermanii*, sont fortement utilisées dans les secteurs de l'agroalimentaire et des biotechnologies. Les connaissances sur la physiologie complexe de cette bactérie peu ordinaire sont faibles et parcellaires, contrairement aux bactéries lactiques dont la physiologie a été largement étudiée depuis de nombreuses années. Dans cet article, nous présentons les avancées de notre projet de séquençage du génome complet de *P. shermanii* : 93% du génome (en 426 contigs) est déjà disponible avec un faible taux global de recouvrement. Notre stratégie d'analyse chevauche le procédé de finition du séquençage et l'améliore. Nous avons mis en place ContigBrowser, un système d'analyse bio-informatique dédié à l'annotation de génome incomplet et à la gestion des données. L'annotation nous a permis de mettre en évidence 2611 protéines putatives (données mai 2003). L'analyse automatique des catégories fonctionnelles de protéines d'intérêt a été soumise à un processus d'expertise par annotation manuelle. L'intégration au sein d'une base de connaissances de l'analyse de notre séquence génomique et des approches protéomiques et métabolomiques, déjà développées au laboratoire, devrait fournir une vision globale des enzymes et des voies métaboliques des bactéries propioniques laitières. Cette vue intégrée du métabolisme pourrait faciliter l'exploration de potentialités significatives des bactéries propioniques laitières. Cette base de connaissances devrait aussi permettre d'explorer de nouvelles stratégies plus efficaces pour améliorer l'utilisation de cette bactérie dans les procédures de fabrication ou les procédés industriels actuels.

***Propionibacterium* / séquençage du génome complet / génomique / annotation de génome / bioinformatique**

1. INTRODUCTION

Dairy propionibacteria play an important role in Swiss-type cheese manufacturing. Having no function in the early stages of cheese manufacture, they grow during the ripening, fermenting mainly lactate and, to a lesser extent, aspartate, to propionate, acetate, succinate and CO_2 (responsible for eyes formation) as major metabolic end-products. Propionic and acetic acids contribute to the nutty and sweet flavour of the cheese (for a review see [5–7]).

Another economically important biotechnological use of propionibacteria is their ability to synthesise high levels of food-grade metabolites such as porphyrins (mainly vitamin B_{12}) and of propionic acid (a natural anti-fungal agent used in packaged bakery products) from cheese-whey or dairy by-products.

For two decades, propionibacteria have had an increasing application for diet supplementation based on their nutraceutical and probiotic properties for animals and humans. They are also added as starter to silage in order to increase its stability.

Propionibacteria are Gram-positive, non-motile, non-sporing, anaerobic to aerotolerant, pleïomorphic and mesophilic bacteria belonging to *Actinobacteria*. The metabolism of these chemoorganotrophs is characterised by their ability to ferment carbohydrates, polyols and organic acids such as lactic acid. Propionibacteria are slow-growing, particularly on solid medium, and their optimal growth rate is between 30 and 37 °C.

Propionibacterium freudenreichii differs from the other dairy propionibacterial species by a high catalase activity and is divided into two subspecies: *P. freudenreichii* subsp. *freudenreichii* (*P. freudenreichii*), harbouring a nitrate reductase activity, but unable to ferment lactose, and *P. freudenreichii* subsp. *shermanii* (*P. shermanii*) lacking nitrate reductase activity, but able to ferment lactose.

Propionibacterial genomes have a high GC content within the range of 53% to 68%. Depending on the species of dairy propionibacteria examined so far, the genome size was estimated to be 1.6 to 3.2 Mb [27]. Genome size varies up to 30% within *P. freudenreichii* species [16, 27].

The genetics of propionibacteria have not yet been extensively investigated. So at the time we started our whole genome sequencing project (November 2000), only scarce genomic data were available: only 30 protein encoding genes were known [34].

Nevertheless, studies on genetic features (antibiotic resistance, plasmids, bacteriophages and genes) are underway and molecular tools (transformation procedures and cloning shuttle vectors) are already under development (for a review see [34]).

Two propionibacterial strains of industrial interest, namely *Propionibacterium acnes* (2.8 Mb, the major microbial agent involved in human acne disease) and *P. freudenreichii* type-strain ATCC 6207 (2.6 Mb), were sequenced by private companies in 2001 [19]. However, these genomes are not publicly available.

The present study reports the progress of the *P. shermanii* genome sequencing project, our annotation strategy and tools, and some of the first bioinformatic analysis of the unfinished genome.

2. MATERIALS AND METHODS

2.1. Organism

Type-strain *P. shermanii* ATCC 9614 (corresponding to the strain CIP 103027 from the Insitut Pasteur collection) was used for the chromosome sequencing. Its genome size is 2.7 ± 0.1 Mb and the strain has no plasmid.

2.2. Library construction

Total genomic DNA from *P. shermanii* was extracted according to the method described by de Carvalho et al. [9] with the following modifications: (i) bacterial cells were harvested at the beginning of the stationary phase and washed with sterile distilled water prior to a pellet freezing and thawing cycle; and (ii) a RNase A treatment was added after the lysis step.

The shotgun library construction followed the *Bst*XI procedure previously described by Dudez et al. [11], which resulted in 9516 recombinant clones with an average insert size of 1.5 kb (varying from 0.8 to 3.8 kb).

2.3. Sequencing and genome assembly strategies

The sequencing project is based on the whole genome sequencing and assembly approach [14]. We used an alternative two-phase sequencing strategy.

In the first stages of the project, all of the 9516 recombinant clones from the shotgun library were sequenced on their forward side. The obtained sequencing reads were assembled (or aligned) in contiguous fragments constructed from many clone sequences called contigs. Then 1615 clones flanking the ends of the obtained contigs were sequenced on their reverse side for extension. This phase corresponds to a typical random whole genome sequencing and assembly approach.

The second phase corresponds to the processes of: (i) further gap closure and contig extension (called finishing), and (ii) obtaining a consensus sequence with higher quality and/or redundancy (called polishing). For this stage we used a rather directed approach. We performed additional reverse sequencing reactions for 1797 recombinant clones flanking the ends of contigs. Suitable clones carrying large inserts were selected and sequenced by 973 internal walk reads in order to close some gaps. The redundancy was kept low by selecting primers that corresponded to the ends of the extended contigs.

Currently (May 2003), 857 additional reads (540 reverse reads and 317 internal walk reads) assuring further gap closure and higher redundancy are underway.

2.4. Bioinformatics

All bioinformatic treatments were performed locally on a bi-processor LINUX workstation.

Whole genome sequence assemblies and manual finishing and polishing were

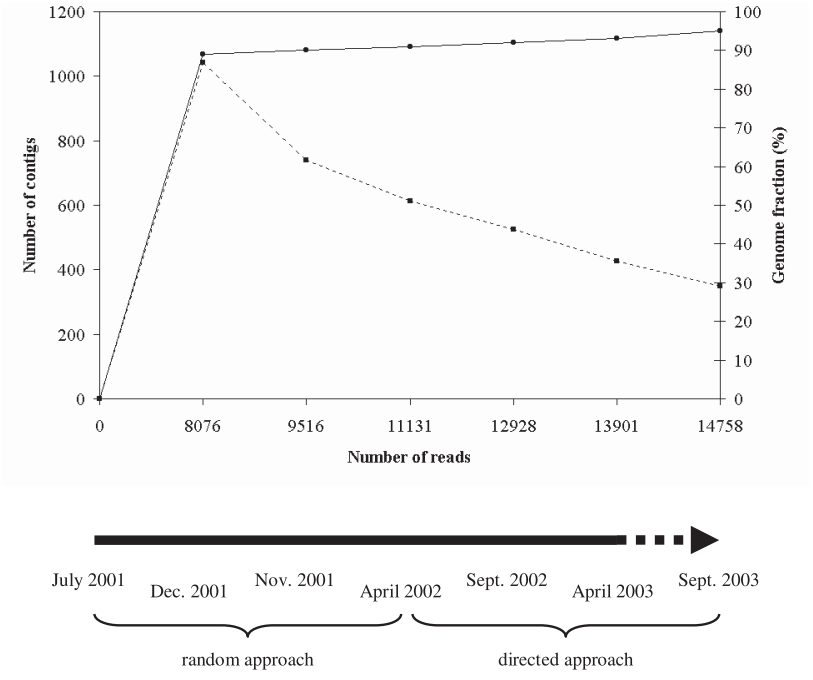

Figure 1. Progress of the sequencing project. The full line corresponds to the increasing genome fraction, while the dashed one is the decreasing number of contigs. The arrow corresponds to the progress of the sequencing project in time.

carried out with the Phred/Phrap/Consed software package [17].

Massive TBLASTN [1] amino-acid sequence comparisons for the directed annotation approach were run manually or in automated mode with a specially-designed Perl script (www.perl.org).

Automated annotation of the *P. shermanii* gapped genome was performed with ContigBrowser, a tool for genome analysis developed in our laboratory and referred to as a pipeline. This pipeline is written in the Perl script language and relies on the MySQL Relational DataBase Management System (www.mysql.com).

3. RESULTS AND DISCUSSION

3.1. Genome sequencing progress

In comparison with fully random sequencing approaches [13], our two-step strategy allowed us to considerably lower the number of sequencing reactions and the cost of the project.

The sequencing of 89% of the *P. shermanii* chromosome was achieved by a classical random shotgun strategy. Successive steps of the directed sequencing strategy, allowing us to lower two-fold the number of contigs and achieve 93% of the genome, followed then. The influence of the growing number of sequencing reads and the chosen strategy on the decreasing number of contigs and the progress of our project is represented in Figure 1.

This combined approach allowed us to establish 93% of the *P. shermanii* chromosome sequence with a low overall sequencing redundancy (3×). This genome fraction is assembled in 426 contigs (genome fragments varying from 1 to 52 kb). With this semi-random, semi-directed strategy,

a sequencing efficacy of about 95% was assured with only 13901 sequencing runs.

However, for the moment, the average sequencing error rate of the project remains relatively high: 40 errors per 10 kb.

Nowadays, the use of fully random sequencing strategy remains a problem in the case of large bacterial genomes. The assembly step is often extremely limited by the presence of numerous repeated elements (ribosomal operons, insertion sequences (IS), and prophages) which may cause false assembly [14]. There were no repetitive sequences from *Propionibacterium* available in public databases at the beginning of our project; thus, we could not exclude them from the initial whole genome assembly. In order to avoid eventual errors, from the first assembly, genome regions containing such repeated sequences were detected and tagged. Those contigs will be assembled separately and finished individually before their re-integration into the final assembly.

Preliminary bioinformatic analysis of the actual genomic draft of *P. shermanii* showed a considerable amount of sequences relative to IS-like elements. It seems that the *P. freudenreichii* genome (sequenced by an industrial consortium) also contains a significant number of repetitive sequences [19].

We are presently attempting to close the remaining gaps between the 426 contigs by different approaches. About 189 of the contigs are organised, ordered and oriented in 73 clearly distinguishable and non-ambiguous chains. This means that 44% of the currently existing gaps between contigs are relatively small and the missing sequence could be easily obtained.

Further gap closure and polishing will be performed by: (i) classical PCR reactions followed by sequencing; (ii) additional reverse and internal walking reactions on shotgun clones; and (iii) sequencing of large genome fragments obtained from rare-cutting enzymes separated by pulse-field gel electrophoresis.

3.2. Annotation strategy and tools

Our annotation strategy overlaps the chromosome finishing process, thus improving it by detection of frameshifts, verification of point-mutations, gap closure directed by truncated proteins identified on the extremities of two different contigs, etc. We have applied two different annotation approaches to the unfinished genome of *P. shermanii*. Their combination significantly refines functional assignment of putative proteins and allows accurate in silico metabolic reconstruction.

The first approach is a directed one and consists of comparing the amino-acid sequence of proteins of interest against raw genomic data. The amino-acid sequences were selected from public databases (Swiss-Prot and KEGG) and chosen according to: (i) their involvement in given metabolic pathways we want to reconstruct (KEGG), and/or (ii) their phylogenetic links to *P. shermanii*. Although partially automated, this approach remains rather expert time-consuming.

To manage automatic ab initio annotation of the gapped genome, as a second approach, we have developed a tool called ContigBrowser.

The first step of the pipeline consists of coding sequences (CDSs) detection with a software tool appropriate for GC-rich genomes. This step was performed with FrameD with the *M. tuberculosis* probabilistic model and default parameters [31]. The gapped genome of *P. shermanii* harbours about 2611 large CDSs (including 33% truncated ones). This software is not only well adapted for finding potential CDSs, but it also helped us with detection of truncated ones, thus assisting further steps for contig finishing.

The following step of our pipeline is to assign function to translated CDSs. Sequence comparisons of translated CDSs against local releases of protein databases (Swiss-Prot, TrEMBL and PIR) were run with BLASTP [1]. Candidate proteins from public

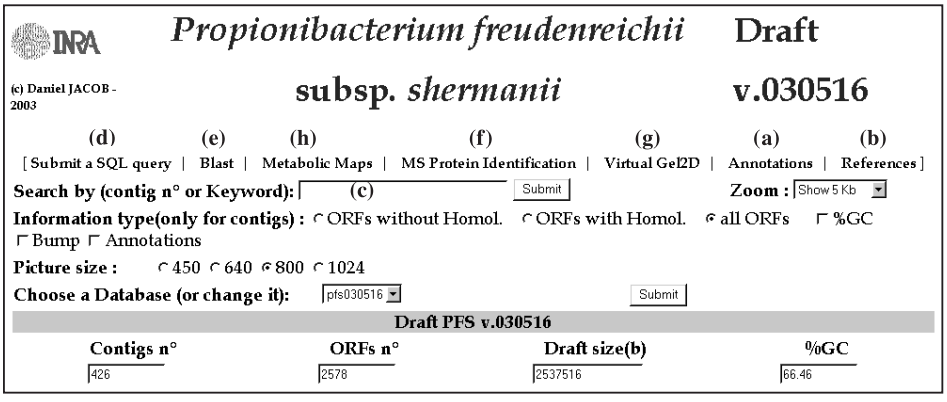

Figure 2. ContigBrowser web interface: **(a)** annotation module; **(b)** module for storage of experimental and bibliographic data; **(c)** information retrieval by keyword or chromosomal localisation (contig); **(d)** MySQL query; **(e)** BLAST searches against genomic data; **(f)** peptide mass fingerprint identification; **(g)** graphical recreation of virtual 2D-gel; and **(h)** KEGG metabolic maps, dynamically tagged with *P. shermanii* genomic data.

databases (E-values lower than 10^{-5}) and corresponding alignments were retained. Then putative propionibacterial proteins were analysed for motifs, domains and functional sites with a standalone version of InterProScan v3.1 [26]. This tool also returns information about protein families and Gene Ontology term annotation. Automatically generated in silico annotation could be further improved by human expertise via a web interface (Fig. 2). When available, annotation (Fig. 2a) of genes of interest could be enriched with related experimental and bibliographic data (Fig. 2b). This cured annotation could then be stored and automatically preserved during drafts' evolution with a home-made CDS-tracer.

Our user-oriented interface for browsing and querying the database is locally accessible through a web server running dynamic HTML pages (Fig. 2). One can easily retrieve information by keywords, chromosomal location (Fig. 2c) or question data tables with a MySQL query (Fig. 2d). This web interface presents some additional features, such as direct BLAST searches against genomic data (Fig. 2e), peptide mass fingerprint identification [28] (Fig. 2f) and home-made script creating a graph of a virtual 2D-gel (Fig. 2g). Visualisation of KEGG metabolic maps, dynamically tagged with *P. shermanii* genomic data (Fig. 2h), represents the first step of in silico metabolic reconstruction.

3.3. First bioinformatic analysis of the unfinished genome

Statistics concerning start and stop codon frequencies of the detected CDSs found in the incomplete chromosome of *P. shermanii* coincide with the translational trends observed in silico in other GC-rich genomes. The AUG and the GUG start and the UGA stop codons are dominant in the *P. shermanii* genome, as in *M. tuberculosis* [29].

Automatic start codon predictions of proteins of interest were corrected manually by checking the RBS presence (level of conservation and distance from the beginning of the largest CDS to the start codon) and multiple sequence alignments with known ortholog proteins (Swiss-Prot).

In the absence of learning samples, the universal mechanism of RBS recognition is base-pairing of the Shine-Dalgarno box and the 3-terminus of the 16S rRNA [32].

Table I. General features of the *P. shermanii* genome.

Chromosomal length	2.7 Mb
Current draft (May 2003)	93% coverage (426 contigs; 3 × redundancy)
GC content	67% (average content)
Putative CDSs	2611 (large CDSs)
Ribosomal RNA operons	2 (1 complete and 1 incomplete)
Transfer tRNAs	45 (including 1 pseudo-tRNA)

The RBS consensus sequence of *Propionibacterium* was established after analysis of the free 3'-end of the two 16S rRNA genes (since multiple copies of 16S rRNA genes present in some bacterial genomes are identical in their anti-Shine-Dalgarno region involved in base-pairing with RBS). The *P. shermanii* RBS consensus (AAAG GAGG) corresponds to the one found in the *Bacillus subtilis* model genome, where it is considered to be a very strong RBS.

The RBS search was performed in an automatic mode during CDS detection. According to the current genomic data, putative RBS were not clearly detected for more than half of the *Propionibacterium* CDSs. In many cases there was no significant base-pairing between the region upstream of the start codons and the 3-terminus of 16S rRNA. This correlates with the observation that GC-rich Gram-positive bacteria tend to possess weak RBS consensus [15].

The average GC content of this bacterium (67%) is among the highest of all bacterial genomes (Tab. I). The base composition is rather homogenous along the chromosome; only intergenic regions and N-terminal gene regions seem to be relatively AT-rich. There are two well-characterised low GC content regions (56%) corresponding to ribosomal RNA genes. But only the GC-plot analysis of the complete *P. shermanii* chromosome will indicate the exact number of integrated prophages and/or horizontal transfer events.

GC- and AT-skews were used to predict the position of replication origin and terminus in the incomplete *P. shermanii* chromosome [22].

A sharp transition between the positive and negative base-skew is observed in the vicinity of the *dnaA* and *dnaN* genes. The locus of replication origin is characterised by high-order synteny conservation: *rnaP-dnaA-oriC-dnaN-recF*. In other genomes this region equally contains four DnaA boxes, also indicating the presence of the origin of replication. However, in the genome of *P. shermanii*, as in the genome of another GC-rich bacterium (*Ralstonia solacearum*) only a single classical consensus DnaA-binding box (TTATCCACA) was observed [30]. A second DnaA box (TTGTCCACA) typical for *Streptomyces* is found upstream of the first one [18]. For the moment, the first nucleotide of the classical consensus DnaA box was arbitrarily chosen as the origin for the nucleotide numbering of the incomplete chromosome. However, six additional strongly degenerated DnaA boxes (containing the minimal consensus TTATC) were detected upstream of the *dnaN* gene, but their role in the replication process is not yet determined.

The approximate position of the replication terminus in the unfinished genome could not be located by the above stated method, so it needs further investigation. No correlation was found between the GC-skew transition from positive to negative values and one putative *dif* site (CAATATATA-TCTC) which is weakly conserved in comparison with the site of the Gram-positive model bacterium *B. subtilis*.

Ribosomal RNA (rRNA) genes were detected and annotated manually by

comparison with ribosomal genes of propionibacterial origin previously sequenced and available in public databases (Y10819, AJ009989 – [8]; X53217 – [3]; AF280074, AF218434, AF218432, AF280075 – [33]). In the current genomic draft, besides the already known ribosomal operon 16S-ITS-5S-ITS-23S (ITS – internally transcribed spacer), we have found another single 16S rRNA gene (Tab. I). The two 16S rRNA genes are 99.6% identical. Despite the high overall genomic GC content of *Propionibacterium*, the two ribosomal loci are characterised by a considerably lower GC ratio – 56%. It appears that, according to the GC-skew method, the complete ribosomal set is apparently located on the leading DNA strand (containing more G than C nucleotides), while the single gene is situated on the lagging strand. The biological signification of the two different locations is yet to be explained.

The 45 transfer RNA (tRNA) genes (including 1 pseudo-tRNA) were detected with the software tRNAScan-SE [23] in bacterial search mode (Tab. I). According to our current genomic data there is no significant close clustering of the tRNA coding genes (4 different tRNAs are co-localised in only 3 cases).

3.4. From functional annotation to metabolic reconstruction

The automatic analysis of some functional categories of proteins of interest was subjected to manual annotation (verification and curation) with the help of our tool, ContigBrowser. Thus 36 genes implicated in stress response [21] and central carbon metabolism [10] were deposited in EMBL and used for in silico metabolic reconstruction of glycolytic pathways [24].

Our preliminary analysis of major functional classes of *Propionibacterium* enzymes involved in energy and central metabolism show an elevated number of paralogs. In *M. tuberculosis* [4] and *Streptomyces* genomes [2] these multigenic families, arising from gene duplication events, lead to extensive functional redundancy. This phenomenon may be related to the common soil origin of all these high GC-rich actinobacteria.

Besides the presence of paralogs we have also evidenced the presence of functional analogs in *P. shermanii* glycolytic pathways. These two observations should be confirmed by further bioinformatic analysis of all enzymatic classes encoded by the complete propionibacterial genome after its completion.

4. CONCLUSIONS AND PERSPECTIVES

Here we report the progress of our low-redundancy whole-genome sequencing project. Up to 93% of the *P. shermanii* genome was obtained with an overall sequencing redundancy close to 3. This genome fraction is assembled in 426 contigs and their number is decreasing constantly.

Being totally supported by public funds, the genome of *P. shermanii* will become publicly available to the scientific community after its completion, contrary to the two competing genome projects (*P. freudenreichii* and *P. acnes*).

Our annotation tool, ContigBrowser, will be employed in the creation of the first knowledge database specialised in propionibacteria (bibliography, physiology, genomics and metabolomics) and its corresponding website will be publicly accessible.

The most important role of whole genome sequencing projects in the post-genomic era is to relate the annotated genome sequence to the physiological reality of the cell. Nowadays, combining careful genome annotation (in silico approach) with in vivo and in vitro methods will make possible the reconstruction of complete metabolic networks of *P. shermanii* [25]. Studies demonstrating approaches of integrative biology are already being used to obtain even quantitative genotype-phenotype relationships of bacterial metabolism [12].

In silico reconstruction of metabolic network could be used to further analyse, interpret and even predict novel metabolic pathways in this microorganism. The exploration of the *P. shermanii* genome is likely to have an impact on the production of improved starter cultures (reproducible growth characteristics, flavour formation and lipolytic activities), the production of bioactive metabolites by current industrial strains (vitamin B_{12} and propionic acid) and on the search for new food-grade bioactive substances or presumed health-promoting factors (bifidogenic factors, bacteriocin, probiotic effects, etc). The whole genome sequence could serve as a good starting point for comparison with other dairy bacteria of industrial importance (such as lactic acid bacteria and bifibobacteria), but also with other pathogenic or commensal propionibacterial strains [20].

Acknowledgements: This project is financed by the Institut National de la Recherche Agronomique (INRA), France. D. Dimova is a postdoctoral researcher supported by INRA, France. G. Meurice is a doctoral student supported by the French Ministry of Education and Research.

REFERENCES

[1] Altschul S.F., Stephen F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucl. Acids Res. 25 (1997) 3389–3402.

[2] Bentley S.D. et al., Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), Nature 417 (2002) 141–147.

[3] Charfreitag O., Stackebrandt E., Inter- and intrageneric relationships of the genus *Propionibacterium* as determined by 16S rRNA sequences, J. Gen. Microbiol. 135 (1989) 2065–2070.

[4] Cole S.T. et al., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.

[5] Conference Proceedings of 1st International Symposium on Propionibacteria, Rennes, France, 17–19 May 1995, Lait 75 (1995) 305–487.

[6] Conference Proceedings of 2nd International Symposium on Propionibacteria, Cork, Ireland, 25–27 June 1998, Lait 79 (1999) 1–200.

[7] Conference Proceedings of 3rd International Symposium on Propionibacteria, Zurich, Switzerland, 8–11 July 2001, Lait 82 (2002) 1–150.

[8] Dasen G.H., Smutny J., Teuber M., Meile L., Classification and identification of propionibacteria based on ribosomal RNA genes and PCR, Biochem. J. 21 (1998) 251–259.

[9] de Carvalho A.F., Guézenec S., Gautier M., Grimont P.A., Reclassification of "*Propionibacterium rubrum*" as *P. jensenii*, Res. Microbiol. 146 (1995) 51–58.

[10] Deborde C., Meurice G., Jacob D., Boyaval P., Dimova D., Combining NMR study and genomic analysis to understand glucose metabolism of *Propionibacterium freudenreichii* subsp. *shermanii*, Poster at Metabolic Engineering IV: Applied System Biology, 6–11 October 2002, Tuscany, Italy, 2002.

[11] Dudez A.M., Chaillou S., Hissler L., Stentz R., Champomier-Verges M.C., Alpert C.A., Zagorec M., Physical and genetic map of the *Lactobacillus sakei* 23K chromosome, Microbiology 148 (2002) 421–431.

[12] Edwards J.S., Ibarra R.U., Palsson B.O., *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data, Nat. Biotechnol. 19 (2001) 125–130.

[13] Fleischmann R.D. et al., Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd, Science 269 (1995) 496–512.

[14] Frangeul L., Nelson K.E., Buchrieser C., Danchin A., Glaser P., Kunst F., Cloning and assembly strategies in microbial genome projects, Microbiology 145 (1999) 2625–2634.

[15] Frishman D., Mironov A., Gelfand M., Starts of bacterial genes: estimating the reliability of computer predictions, Gene 234 (1999) 257–265.

[16] Gautier M., Mouchel N., Rouault A., Sanséau P., Determination of genome size of *Propionibacterium* species by pulse-field gel electrophoresis, Lait 72 (1992) 421–426.

[17] Gordon D., Abajian C., Green P., Consed: A graphical tool for sequence finishing, Genome Res. 8 (1998) 195–202.

[18] Jakimowicz D., Majka J., Messer W., Speck C., Fernandez M., Martin M.C., Sanchez J., Schauwecker F., Keller U., Schrempf H., Zakrzewska-Czerwinska J., Structural elements of the *Streptomyces oriC* region and their interactions with the DnaA protein, Microbiology 144 (1998) 1281–1290.

[19] Klaenhammer T. et al., Discovering lactic acid bacteria by genomics, Antonie van Leeuwenhoek 82 (2002) 29–58.

[20] Kuipers O.P., Genomics for food biotechnology: prospects of the use of high-throughput technologies for the improvement of food microorganisms, Curr. Opin. Biotechnol. 10 (1999) 511–516.

[21] Leverrier P., Dimova D., Pichereau V., Auffray Y., Boyaval P., Jan G., Susceptibility and adaptive response to bile salts in *Propionibacterium freudenreichii*: physiological and proteomic analysis, Appl. Environ. Microbiol. 69 (2003) 3809–3818.

[22] Lobry J.R., Asymmetric substitution patterns in the two DNA strands of bacteria, Mol. Biol. Evol. 13 (1996) 660–665.

[23] Lowe T.M., Eddy S.R., tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucl. Acids Res. 25 (1997) 955–964.

[24] Meurice G., Jacob D., Dimova D., Boyaval P., Deborde C., Reconstruction *in silico* des voies glycolytiques selon le génome de *Propionibacterium freudenreichii*, Poster at 12e Réunion du Club des Bactéries Lactiques, « Des gènes … aux écosystèmes », 19–21 May, Aurillac, France, 2003.

[25] Meurice G., Leverrier P., Rouault A., Maillard M.-B., Thierry A., Corre C., Jacob D., Dimova D., Deborde C., Jan G., Combining genomic, proteomic, NMR and physiological study to understand salt stress response in *Propionibacterium freudenreichii*, Poster at 7e Congrès « RMN biologique et les enjeux du vivant », 12–14 March, Clermont-Ferrand, France, 2003.

[26] Mulder N.J. et al., The InterPro Database, 2003 brings increased coverage and new features, Nucl. Acids Res. 31 (2003) 315–318.

[27] Rehberger T.G., Genome analysis of *Propionibacterium freudenreichii* by pulse-field gel electrophoresis, Curr. Microbiol. 27 (1993) 21–25.

[28] Rice P., Longden I., Bleasby A., EMBOSS: The European Molecular Biology Open Software Suite, Trends Genet. 16 (2000) 276–277.

[29] Rocha E.D.C., Danchin A., Viari A., Translation in *Bacillus subtilis*: role and trends of initiation and termination, insights from a genome analysis, Nucl. Acids Res. 27 (1999) 3567–3576.

[30] Salanoubat M. et al., Genome sequence of the plant pathogen *Ralstonia solanacearum*, Nature 415 (2002) 497–502.

[31] Schiex T., Gouzy J., Moisan A., de Oliveira Y., FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy maturated eukaryotic sequences, Nucl. Acids Res. 31 (2003) 3738–3741.

[32] Shine J., Dalgarno L., The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites, Proc. Natl. Acad. Sci. USA 71 (1974) 1342–1346.

[33] Tilsala-Timisjarvi A., Alatossava T., Characterization of the 16S-23S and 23S-5S rRNA intergenic spacer regions of dairy propionibacteria and their identification with species-specific primers by PCR, Int. J. Food Microbiol. 68 (2001) 45–52.

[34] van Luijk N., Stierli M.P., Miescher Schwenninger M., Hervé C., Dasen G., Jore J.P.M., Pouwels P.H., van der Werf M.J., Teuber M., Meile L., Genetics and molecular biology of propionibacteria, Lait 82 (2002) 45–57.