



HAL
open science

**Analyse de la séquence génomique et Etude de
l'adaptation à l'acidité de *L. delbrueckii* ssp. *bulgaricus*
ATCC11842**

Stéphanie Penaud

► **To cite this version:**

Stéphanie Penaud. Analyse de la séquence génomique et Etude de l'adaptation à l'acidité de *L. delbrueckii* ssp. *bulgaricus* ATCC11842. Life Sciences [q-bio]. INAPG (AgroParisTech), 2006. English. NNT : 2006INAP0013 . pastel-00002283

HAL Id: pastel-00002283

<https://pastel.hal.science/pastel-00002283>

Submitted on 3 Apr 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL AGRONOMIQUE
DE PARIS-GRIGNON

THESE

Présentée par :

Stéphanie PENAUD

Pour obtenir le grade de :

Docteur de l'Institut National Agronomique de Paris-Grignon

**Analyse de la séquence génomique et
Etude de l'adaptation à l'acidité
de *L. delbrueckii* ssp. *bulgaricus* ATCC11842**

Soutenue le 10 juillet 2006

Devant le jury composé de :

Mr Jean-Claude Meunier

Président

Mr Jean Guzzo

Rapporteur

Mr Paul Ritzenthaler

Rapporteur

Mme Tamara Smokvina

Examinatrice

Mr Maarten van de Guchte

Examineur

Mme Emmanuelle Maguin

Directrice de thèse

Remerciements

Je tiens tout d'abord à remercier les membres du jury d'avoir accepté d'évaluer mon travail et de l'investissement que cela représente. Merci à Jean-Claude Meunier de m'avoir suivie au cours de ma thèse, à Jean Guzzo et Paul Ritzenthaler d'avoir accepté d'être rapporteurs et à Tamara Smokvina d'avoir accepté d'être examinatrice.

Je remercie Dusko Ehrlich de m'avoir accueillie au sein de l'unité Génétique Microbienne et de m'avoir permis de terminer ma thèse dans de bonnes conditions et Danone pour avoir financé une partie de ce travail.

Je tiens bien sûr à remercier tous les membres de l'équipe « Stress ». Et tout d'abord Emmanuelle Maguin pour m'avoir permis de réaliser cette thèse et pour tous les conseils prodigués au cours de celle-ci. Je suis contente d'être allée au bout... Merci aussi à Maarten, pour tous les moments partagés et tes conseils pour ce manuscrit. Une pensée particulière pour mes deux « sœurs » du labo, Christine et Samira, nos longues discussions sur le labo, la vie et le reste me manquent déjà... N'oubliez jamais que vous êtes toutes les deux exceptionnelles. Annabelle, merci pour le rayon de soleil que tu as apporté ici, et pourtant venant de Lorraine, c'était pas gagné. Ton passage aura été trop court. Merci à Isabelle, pour ta disponibilité et ta bonne humeur. Merci aussi à Rozenn et à Frede de m'avoir accueillie dans le bureau et à Luciana pour ta joie de vivre et ton enthousiasme (c'est toi la prochaine...).

Merci à toute l'unité de Génétique Microbienne, et plus particulièrement à Pierre, Nico, Benjamin, JC et Jean-Michel pour les discussions partagées. Je remercie tous les gens de la prépa pour leur aide logistique si précieuse. Spéciale dédicace à Ludo pour ton soutien. Merci à Jean-Pierre Furet pour son encadrement si précieux pour la PCR quanti. Merci à l'unité MIG pour leur disponibilité et le programme Agmial, et plus particulièrement à Valentin, Robert et Pierre.

Merci à l'équipe Doc'J, vous m'avez si souvent rendu le sourire, Fabien, Jeff, Pat, Benjamin, Béa, Lionel, Tat, les deux David, et les autres, l'expérience aura été si enrichissante grâce à vous. Merci aussi au groupe PPP, Eric, Karine, Cécilia et Mariam notamment pour les bons moments passés. Merci à Lahcen pour les discussions animées.

Je remercie aussi tous ceux qui m'ont accompagnée pendant ces années en dehors du labo et qui ont été là quand j'en avais besoin. Merci à mes parents de m'avoir permis de réaliser mon rêve. J'espère que vous êtes fiers de moi. Fanny, Nath, Mél, votre amitié aura été d'un soutien inestimable et j'essaierai d'être à l'avenir plus disponible...

Et enfin, le plus grand des mercis à ma moitié, pour avoir toujours été là quand j'en avais besoin et pour avoir su comprendre les concessions et les sacrifices. Ton amour, ton soutien, ta patience, ta compréhension et le réconfort que tu m'as apporté auront été irréprochables. Cette thèse est aussi la tienne...

SOMMAIRE

Liste des principales abréviations.....	5
INTRODUCTION.....	7
1. LES BACTÉRIES LACTIQUES, UN GROUPE COMPLEXE ET HÉTÉROGÈNE	9
1.1. DES BACTÉRIES UTILES À L'HOMME	9
1.1.1. Une utilisation millénaire	9
1.1.2. Caractéristiques générales des Bactéries Lactiques.....	10
1.1.3. Utilisation industrielle des bactéries lactiques	10
1.1.4. Des bactéries commensales de l'homme.....	11
1.2. CLASSIFICATION ET TAXONOMIE.....	13
1.3. LES LACTOBACILLES : UN GENRE REPRÉSENTATIF DE LA DIVERSITÉ DES BL	16
1.3.1. Diversité génétique	16
1.3.2. Diversité des habitats.....	17
1.3.3. Effets des lactobacilles sur la santé	18
1.3.4. Le groupe acidophilus.....	22
2. APPORTS DE LA GÉNOMIQUE SUR L'ÉTUDE DES BACTÉRIES LACTIQUES.	24
2.1. GÉNOMIQUE ET GÉNOMIQUE COMPARATIVE DES CAPACITÉS D'ADAPTATION DES SOUCHES.....	24
2.1.1. Informations issues d'un génome.....	24
2.1.1.1. Structure du génome	24
2.1.1.2. Capacités métaboliques	26
2.1.2. Comparaison de génomes	27
2.1.3. Génome minimal et ancêtre commun.....	31
2.2. RECONSTRUCTION DES RÉSEAUX DE RÉGULATION	34
2.2.1. Généralités sur les systèmes de régulation transcriptionnelle.....	34
2.2.2. Identification de régulons	37
3. LES RÉPONSES AUX STRESS	38
3.1. RÉGULATION DES MÉCANISMES DE RÉSISTANCE AU STRESS	38
3.1.1. La réponse générale au stress.....	38
3.1.2. Le stimulon de réponse au stress thermique : exemple de régulons coordonnés.....	39
3.1.3. Les BL et la régulation des réponses aux stress.....	41
3.2. LE STRESS ACIDE.....	43
3.2.1. Effets de l'acide lactique sur les bactéries.....	43
3.2.2. Mécanismes de résistance au stress acide des BL	43
3.2.2.1. Maintien de la force motrice.....	44
3.2.2.2. Production de composés alcalins	49

3.2.2.3. Modifications de l'enveloppe cellulaire.....	51
3.2.2.4. Prise en charge des protéines dénaturées	52
3.2.2.5. Réparations de l'ADN.....	53
4. L. DELBRUECKII SSP. BULGARICUS.....	55
RESULTATS	57
PARTIE I : ANALYSE DU GENOME DE L. BULGARICUS ATCC11842.....	57
1. INTRODUCTION	59
2. ANNOTATION DU GÉNOME.....	59
2.1. AGMIAL, PRÉSENTATION ET ÉVOLUTION DE LA PLATEFORME D'ANNOTATION.....	59
2.1.1. Présentation d'Agmial.....	60
2.1.2. Validation des annotations	64
2.2. COMPARAISON ERGO/AGMIAL.....	69
2.2.1. Prédiction de gènes	69
2.2.2. Annotation automatique	69
2.2.3. Reconstructions métaboliques	70
3. ANALYSE DES DONNÉES GÉNOMIQUES	70
3.1. CARACTÉRISTIQUES GÉNÉRALES DES CDS	70
3.2. CAPACITÉS MÉTABOLIQUES	73
3.2.1. Métabolisme des acides aminés.....	73
3.2.2. Métabolisme des bases azotées.....	74
3.3. LES SYSTÈMES DE RÉGULATION TRANSCRIPTIONNELLE	74
3.3.1. Les facteurs sigma	74
3.3.2. Les systèmes à deux composants	77
3.3.3. Les régulateurs	78
3.4. PROPRIÉTÉS PROBIOTIQUES ET SURVIE DANS LE TRACTUS DIGESTIF	79
3.5. LES FONCTIONS IMPLIQUÉES DANS LES RÉPONSES AU STRESS	81
3.5.1. Le stress thermique.....	81
3.5.1.1. Le stress hyperthermique	81
3.5.1.2. Le stress hypothermique	84
3.5.2. Le stress oxydant	85
3.5.3. Réparation des lésions ADN : le système SOS.....	87
3.5.4. Le stress osmotique.....	88
3.5.5. Le stress acide	90
3.6. CONCLUSION	91

4. THE COMPLETE GENOME SEQUENCE OF LACTOBACILLUS BULGARICUS REVEALS AN EXTENSIVE AND ONGOING REDUCTIVE EVOLUTION.....	92
ANNEXES.....	125

PARTIE II : ETUDE DE L'ADAPTATION A L'ACIDITE147

1. AVANT-PROPOS.....149

2. ETUDES EN MILIEU CHIMIQUEMENT DÉFINI150

2.1. EXPRESSION DES RÉGULATEURS EN CONDITION D'ADAPTATION À L'ACIDE LACTIQUE.....	150
2.1.1. <i>Introduction</i>	150
2.1.2. <i>Mise au point des conditions expérimentales</i>	151
2.1.2.1. Conditions d'adaptation en MPL et préparation des échantillons.....	151
2.1.2.2. PCR quantitative et gène de référence	152
2.1.3. <i>Résultats</i>	153
2.2. ETUDE DE L'EXPRESSION DES SYSTÈMES DE TRANSPORT D'IONS.....	159
2.2.1. <i>Détermination du gène de référence</i>	159
2.2.2. <i>Expression des transporteurs d'ions lors de l'ATR</i>	160
2.2.3. <i>Résultats complémentaires</i>	185
2.2.3.1. Stress et adaptation croisés	185
2.2.3.2. Mécanismes de régulation par atténuation.....	185
2.3. DISCUSSION	188

3. ETUDE EN MILIEU COMPLEXE191

3.1. INTRODUCTION	191
3.2. MATÉRIEL ET MÉTHODES	191
3.3. RÉSULTATS	196
3.3.1. <i>Résistance à l'acidité en MRS</i>	196
3.3.2. <i>Adaptation</i>	197
3.3.3. <i>Protéomique</i>	197
3.3.4. <i>Etude transcriptionnelle</i>	200
3.4. DISCUSSION	202
3.4.1. <i>ATR et stimulon de réponse au choc thermique</i>	202
3.4.2. <i>ATR et métabolisme carboné</i>	203
3.4.3. <i>ATR et métabolisme des acides gras</i>	204

DISCUSSION GENERALE / PERSPECTIVES	207
1. ADAPTATION ET DÉRIVE GÉNÉTIQUE	209
1.1. EVOLUTION DU GÉNOME.....	209
1.2. PSEUDOGÈNES ET ADAPTATION	212
2. RÉPONSE ADAPTATIVE À L'ACIDITÉ DE L. BULGARICUS	213
2.1. RÉGULATION DE L'ATR	214
2.2. ATR ET RÉGULON DU CHOC THERMIQUE	215
Bibliographie.....	217
Publications annexes	243

LISTE DES PRINCIPALES ABREVIATIONS

2DE : Two-dimensional electrophoresis

A : Absorbance

ABC : ATP Binding Cassette

ADNc : ADN complémentaire

ATR : Acid Tolerance Response

BL : Bactéries Lactiques

CDS : Coding Sequence

MCD : Milieu Chimiquement défini

pb : paire de bases

PCR : Polymerase Chain Reaction

pI : Point Isoélectrique

PM : Poids Moléculaire

RBS : Ribosome Binding Site

RT : Reverse Transcription

ssp. : sous-espèce

Ufc : Unité formant colonie

Introduction

1. Les bactéries lactiques, un groupe complexe et hétérogène

1.1. Des bactéries utiles à l'Homme

Les bactéries lactiques sont utilisées depuis des millénaires dans l'alimentation humaine. Après un bref rappel historique sur l'utilisation des produits fermentés et notamment des produits laitiers au cours de l'histoire humaine, nous verrons quelles sont leurs caractéristiques principales et comment elles sont utilisées industriellement de nos jours. Nous aborderons ensuite la relation entre les bactéries commensales et l'hôte, point qui a pris de l'importance dans le domaine des bactéries lactiques avec l'utilisation de bactéries probiotiques en alimentation humaine et la multiplication des produits probiotiques commerciaux.

1.1.1. Une utilisation millénaire

L'utilisation de la fermentation par l'Homme remonte à des temps très anciens. Les premiers produits fermentés ont certainement été obtenus par acidification spontanée des jus végétaux (vins, bières...) ou par contamination naturelle du lait (yaourts, fromages...). Les premières preuves de l'existence des produits laitiers fermentés remontent à 8000 ans avant JC dans le croissant fertile au Moyen Orient (plaines du Nil, du Jourdain, de l'Euphrate et du Tigre), époque où les végétaux et les animaux sont domestiqués (Fox, 1993). La fermentation des végétaux (vins, bières) et la production de levain apparaissent entre 4000 et 2000 avant JC chez les Egyptiens. La fermentation est réalisée à partir de différents types d'aliments : des végétaux (concombres, betteraves, dattes, jus de fruits, soja, etc...), des produits animaux (viande, lait) ou du poisson. Elle permet de conserver les aliments mais aussi de leur donner une saveur différente du produit original.

Il faudra attendre Pasteur et ses travaux sur la fermentation en 1857 (Pasteur, 1857) pour établir un lien entre la fermentation lactique et les bactéries. La première culture bactérienne pure sera d'ailleurs une culture de *Lactococcus lactis* obtenue et décrite par Joseph Lister en 1873 (Lister, 1873). Metchnikoff isole en 1904 le « bacille bulgare » (*Lactobacillus delbrueckii* ssp. *bulgaricus*) présent dans le yaourt. Il étudie les propriétés acidifiantes des bactéries du yaourt et il développera l'idée que les bactéries contenues dans les laits fermentés ont un effet bénéfique

sur la santé (Metchnikoff, 1907). Il plaidera en faveur de l'introduction de produits laitiers fermentés dans le régime alimentaire et en 1905, les premières entreprises fabricant du yaourt à partir des souches de l'Institut Pasteur voient le jour (Bibel, 1988).

Depuis 1960, la production industrielle de yaourt connaît un développement international considérable et le yaourt est devenu aujourd'hui un produit de consommation courante en Occident. En 2003, la consommation de yaourts et de laits fermentés s'est élevée en France à 1,3 millions de tonnes, soit 21,9 kg par an et par habitant (contre 17,7 en 1994, source : www.cniel.com) et le chiffre d'affaire représente 2097 millions d'euros (source : www.cidilait.com).

1.1.2. Caractéristiques générales des Bactéries Lactiques

La première définition de bactéries lactiques (BL), basée sur la capacité des bactéries de fermenter et de coaguler le lait, englobait les bactéries coliformes et lactiques. En 1901, Beijerinck observe que les lactobacilles sont des bactéries à Gram positif, ce qui séparera définitivement les bactéries lactiques (à Gram positif) des bactéries coliformes (Stiles et Holzapfel, 1997).

Les bactéries lactiques sont donc des bactéries à Gram positif qui convertissent le pyruvate en acide lactique pour régénérer le NAD^+ utilisé dans la glycolyse. A quelques exceptions près, elles partagent les caractéristiques suivantes : elles sont généralement immobiles, asporulées, anaérobies mais aérotolérantes. Pour se développer, elles ont besoin de sources de carbone organique (glucides fermentescibles) et de nombreuses BL ont des exigences nutritionnelles complexes en ce qui concerne les acides aminés ou les peptides, les vitamines et les acides gras (Prescott *et al.*, 1999).

Les bactéries lactiques peuvent avoir un métabolisme homofermentaire (plus de 90% des produits de fermentation est de l'acide lactique), hétérofermentaire facultatif (elles produisent de l'acide lactique ou de l'acide lactique et de l'acide acétique) ou hétérofermentaire strict (elles produisent, en plus de l'acide lactique, de l'acide acétique ou de l'éthanol et du CO_2) (Vandamme *et al.*, 1996).

1.1.3. Utilisation industrielle des bactéries lactiques

Les bactéries lactiques (BL) sont très utilisées dans l'industrie agroalimentaire. Elles peuvent fermenter une grande variété de substrats (Tableau 1).

Tableau 1 : Principaux produits issus de la fermentation des bactéries lactiques

Genre	Substrat	Exemples de produits
<i>Bifidobacterium</i>	lait	laits fermentés
<i>Lactobacillus</i>	lait	yaourts, laits fermentés, kéfirs, fromages
	viande	saucissons secs, jambons secs
	végétaux	choucroute, olives, "yaourts" au lait de soja
	céréales	pain au levain, bières
<i>Lactococcus</i>	lait	fromages, kéfirs
<i>Leuconostoc</i>	végétaux	choucroute, olives, vin
	lait	fromages, kéfirs
<i>Pediococcus</i>	végétaux	choucroute
	viande	saucisses semi-séchées
<i>Oenococcus</i>	végétaux	vin
<i>Streptococcus</i>	lait	yaourts, laits fermentés, fromages

Les BL ont plusieurs rôles dans la production de produits fermentés. Tout d'abord, les BL vont permettre de changer la saveur de l'aliment et sa texture. Ces changements sont dus notamment à l'acide lactique produit au cours de la croissance. D'autre part, les BL produisent des peptides et des molécules comme l'acétoïne, l'acétaldéhyde, le diacétyle ou l'éthanol qui sont importants pour la flaveur des aliments. Un autre rôle des BL est d'inhiber le développement de la flore bactérienne indésirable qu'elle soit pathogène ou d'altération. Cette inhibition passe par deux aspects : l'acidification qui inhibe la croissance des bactéries peu résistantes à un bas pH, et la synthèse de bactériocines, molécules bactéricides dont les spectres d'action sont variables : des bactéries à Gram + (nisine par exemple ; Klaenhammer, 1988) à l'espèce productrice (caséicine 80 par exemple ; Rammelsberg *et al.*, 1990).

Certaines BL comme *L. helveticus* sont utilisées pour produire industriellement de l'acide lactique employé comme additif en alimentation et dans les produits cosmétiques ou pharmaceutiques. L'acide lactique est aussi transformé en acide polylactique, polymère bio-résorbable utilisé dans la fabrication d'implants pour la chirurgie osseuse ou pour la fabrication de films plastiques biodégradables. *Leuconostoc mesenteroides* est utilisé pour la production de dextran, médicament utilisé dans les cas d'hypovolémie sanguine.

1.1.4. Des bactéries commensales de l'homme

Si les BL sont importantes pour l'alimentation humaine, elles le sont aussi pour sa santé. En effet, des BL ont colonisé un grand nombre de niches écologiques notamment le tractus digestif. Le tractus digestif humain est l'écosystème présentant la plus grande densité cellulaire décrite jusqu'à aujourd'hui et le nombre de bactéries présentes dans le tractus gastro-intestinal ($\sim 10^{15}$)

est supérieur au nombre total de cellules humaines constituant l'organisme (Backhed *et al.*, 2005). Cette microflore évolue au cours de la vie, notamment dans la première année suivant la naissance. Les premières bactéries à coloniser le tube digestif sont des bactéries coliformes et des BL des genres *Lactobacillus* et *Streptococcus* (Isolauri *et al.*, 2004). Cette colonisation précoce permettrait d'obtenir des conditions plus favorables à la colonisation par la flore définitive, composée en majorité (plus de 99%) de bactéries anaérobies strictes comme les bactéries du genre *Bacteroides* et *Bifidobacterium* (McCracken et Lorenz, 2001). On estime que seuls 20 à 40% des bactéries du colon sont cultivables et ont été caractérisées. L'identification et la quantification de ces bactéries sont en cours notamment par des études faisant appel au ARNr 16S (Servin, 2004). Les BL que l'on retrouve dans le colon à l'âge adulte appartiennent aux genres *Bifidobacterium*, *Lactobacillus* et *Enterococcus* et font partie de la flore minoritaire (Berg, 1996).

Les rôles principaux de la flore digestive portent sur des aspects métaboliques, trophiques et immunitaires (Guarner et Malagelada, 2003). La flore intestinale permet de fermenter des aliments non digérables par l'homme ainsi que le mucus produit par l'épithélium intestinal, ce qui lui permet de produire des acides gras à courtes chaînes carbonées. Ces acides gras (notamment l'acétate, le propionate et le butyrate) sont importants pour l'hôte et sont utilisés comme source d'énergie (Cummings et Macfarlane, 1997). La microflore intestinale joue aussi un rôle dans la synthèse de vitamines (Hill, 1997) et dans l'absorption de certains ions comme le calcium, le magnésium et le fer (Miyazawa *et al.*, 1996). Les bactéries résidentes du tube digestif ont un rôle trophique sur l'épithélium intestinal. La différenciation du tube digestif et la prolifération des cellules épithéliales sont dépendantes de l'interaction avec la microflore. En effet, les animaux axéniques (sans flore intestinale) présentent un intestin immature avec des cryptes moins nombreuses et contenant moins de cellules que les animaux conventionnels (Alam *et al.*, 1994).

Les bactéries du tube digestif protégeraient l'hôte des infections bactériennes à deux niveaux : elles stimuleraient le système immunitaire de l'hôte et le protégeraient des pathogènes par « l'effet barrière ». Les tissus lymphoïdes associés à l'intestin sont les tissus contenant le plus de cellules immunocompétentes de tout l'organisme. Cette concentration de cellules immunes est dépendante de la présence d'une flore intestinale (Guarner et Malagelada, 2003). La stimulation immunitaire touche aussi bien la réponse acquise que la réponse innée de l'hôte. Les mécanismes immunitaires mis en jeu dans l'interaction intestin-microflore ne seront pas détaillés ici (pour revue : Servin, 2004). La fonction protectrice de la flore sur les bactéries pathogènes porte sur

deux aspects : la compétition pour l'habitat et la production d'agents anti-microbiens (Servin, 2004). Cette protection impliquerait notamment les bactéries lactiques commensales et a été étudiée pour certaines espèces des genres *Bifidobacterium* et *Lactobacillus* (les études portant sur les lactobacilles seront détaillées § 1.3.3).

1.2. Classification et taxonomie

La première classification des bactéries lactiques a été établie en 1919 par Orla-Jensen sur divers critères morphologiques et physiologiques (activités catalase et nitrite réductase, type de fermentation) (Stiles et Holzapfel, 1997). Les critères phénotypiques permettant la classification des bactéries se sont ensuite étendus à la composition de la paroi, le type d'acides gras cellulaires, le type de quinones (accepteur d'électrons). Cependant ces méthodes phénotypiques ne rendent pas compte des relations phylogénétiques entre les groupes. En 1977, Woese et Fox introduisent la phylogénie moléculaire basée sur la séquence des ARN ribosomiques. Cette méthode va révolutionner la taxonomie des bactéries et la classification des BL va être profondément modifiée. D'autres méthodes génotypiques (basés sur les acides nucléiques) sont aussi utilisées en classification, comme le pourcentage en GC ou l'hybridation ADN:ADN. L'approche consistant à prendre en compte les méthodes phénotypiques et génotypiques s'appelle la taxonomie polyphasique (Vandamme *et al.*, 1996). La technique de MLST (pour Multi Locus Sequence Typing), basée sur la divergence nucléique de gènes de ménage, est utilisée pour la classification des BL pathogènes (notamment streptocoques ; Enright et Spratt, 1999). Cependant cette technique n'a pas été utilisée pour d'autres BL.

Le groupe des BL ne peut pas être considéré comme un groupe phylogénétique. Elles appartiennent toutes au groupe des bactéries à Gram positif, mais si la plupart d'entre elles appartiennent au groupe des bactéries à Gram positif à bas G+C (phylum des Firmicutes), le genre *Bifidobacterium* appartient au groupe des bactéries à Gram positif à haut G+C (phylum des Actinomycètes) (Figure 1). Si les bifidobactéries sont phylogénétiquement éloignées des BL *sensu stricto* (BL à bas G+C), elles sont tout de même traditionnellement incluses dans les BL car elles partagent certaines caractéristiques avec elles (elles produisent de l'acide lactique et sont utilisées dans les laits fermentés ; Holzapfel *et al.*, 2001).

Parmi les bactéries lactiques *sensu stricto*, on peut distinguer quatre groupes phylogénétiques principaux. Les genres *Enterococcus*, *Melissococcus*, *Tetragenococcus*, *Vagococcus*, *Carnobacterium*, *Lactosphaera* et *Aerococcus* forment un premier groupe (Stiles et

Holzapfel, 1997). Dans ce groupe, l'effort de séquençage des génomes a porté sur les espèces pathogènes opportunistes de l'homme, *Enterococcus faecalis* et *E. faecium* (Tableau 2).

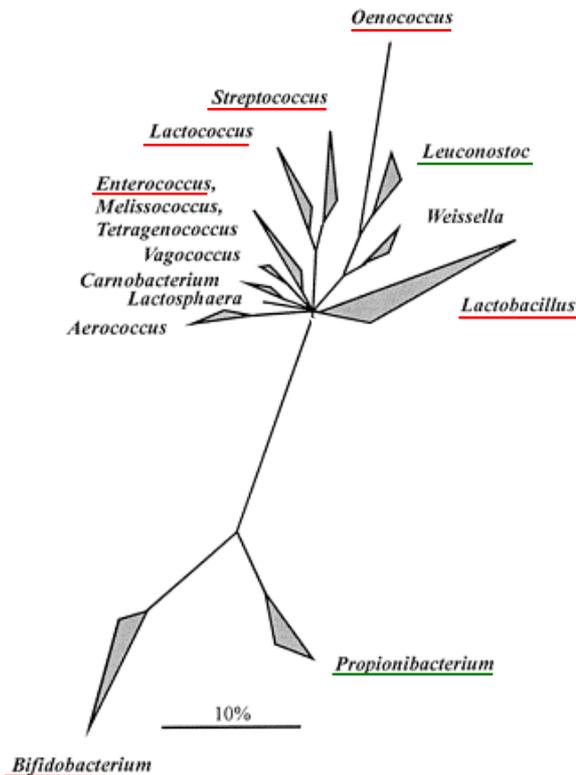


Figure 1 : Distances phylogénétiques entre les genres principaux constituant les BL basées sur les séquences des ARNr 16S.

Les bactéries à bas G+C (en haut) sont phylogénétiquement éloignées des bactéries à haut G+C (en bas). Sont soulignés en rouge les genres pour lesquels la séquence d'au moins un génome est disponible, en vert ceux pour lesquels au moins un génome est en cours de séquençage.

La barre indique une divergence de séquence estimée de 10%.

Adapté de Schleifer et Ludwig, 1995

Les genres *Streptococcus* et *Lactococcus* n'ont été séparés qu'en 1985 avec l'apparition du genre *Lactococcus* (Schleifer *et al.*, 1985) et sont phylogénétiquement proches. *Lactococcus lactis* est une bactérie très utilisée en fermentation fromagère. Le premier génome de BL disponible a été celui d'une souche de *Lactococcus lactis* ssp. *lactis* en 2001 (Bolotin *et al.*, 2001). Dans le genre *Streptococcus*, la seule espèce utilisée comme ferment dans l'industrie alimentaire est *S. thermophilus*, notamment pour la production de yaourt avec *L. bulgaricus*. Ce genre contient un grand nombre de bactéries plus ou moins pathogènes (*S. pneumoniae*, *S. pyogenes*, *S. mutans*, *S. agalactiae*...) pour lesquels 13 séquences génomiques sont d'ores et déjà

disponibles (non présentés dans le tableau 2, voir <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>).

Tableau 2 : BL sensu stricto dont les génomes sont séquencés ou en cours de séquençage

Genre	Espèce	Souche	Habitat d'origine	Référence
<i>Enterococcus</i>	<i>faecalis</i>	V583	Sang humain	Paulsen <i>et al.</i> 2005
<i>Lactococcus</i>	<i>lactis</i>	<i>lactis</i> IL1403	Dérivé d'un ferment fromager	Bolotin <i>et al.</i> 2001
<i>Streptococcus</i>	<i>thermophilus</i>	CNRZ1066	Yaourt	Bolotin <i>et al.</i> 2004
		LMG18311	Yaourt	
<i>Oenococcus</i>	<i>oeni</i>	PSU-1	Vin	Mills <i>et al.</i> 2005
<i>Lactobacillus</i>	<i>acidophilus</i>	NCFM	Tractus digestif humain	Altermann <i>et al.</i> 2005
	<i>johnsonii</i>	NCC 533	? Tractus digestif humain	Pridmore <i>et al.</i> 2004
	<i>plantarum</i>	WCFS1	Salive humaine	Kleerebezem <i>et al.</i> 2003
	<i>sakei</i>	<i>sakei</i> 23K	Saucisse	Chaillou <i>et al.</i> 2005
	<i>salivarius</i>	<i>salivarius</i> UCC118	Tractus digestif humain	Claesson <i>et al.</i> 2006
Genre	Espèce	souche	Habitat d'origine	Groupe
<i>Enterococcus</i>	<i>faecium</i>	DO	Humain	JGI
<i>Lactococcus</i>	<i>lactis</i>	<i>cremoris</i> SK11	Fromage	JGI / LABGC
		<i>cremoris</i> MG1363	Fromage	Univ. de Groningen, Cork et Bielefeld/IFR
		<i>cremoris</i> QA5	Produits laitiers	INRA
		<i>cremoris</i> KF147 & KF282	Plantes	NIZO Food Research
<i>Streptococcus</i>	<i>thermophilus</i>	LMD-9	Yaourt et mozzarella	JGI / LABGC
		DN-001 147	Produits laitiers	Danone - INRA
<i>Leuconostoc</i>	<i>citreum</i>	KM20	Végétaux fermentés	KRIBB
		<i>mesenteroides</i> ATCC 8293	Végétaux fermentés	JGI / LABGC
<i>Oenococcus</i>	<i>oeni</i>	IOEB84.13	Vin	INRA
<i>Lactobacillus</i>	<i>brevis</i>	ATCC 367	? Végétaux fermentés	JGI / LABGC
		ATCC 334	Fromage	JGI / LABGC
	<i>casei</i>	Shirota	?	Yakult
		DN-114 001	Lait fermenté	Danone - INRA
		BL23	? Lait fermenté	INRA/CNRS - Université de Caen
	<i>delbrueckii</i>	<i>bulgaricus</i> ATCCBAA-365	Yaourt	JGI / LABGC
		<i>bulgaricus</i> DN 100-107	? Yaourt	Danone
		<i>bulgaricus</i> ATCC 11842	Yaourt	INRA/Génoscope
	<i>gasseri</i>	ATCC 33323	? Tractus digestif humain	JGI / LABGC
	<i>helveticus</i>	CNRZ32	? Laits fermentés et fromage	Université du Wisconsin
		CPC4571	Fromage	Université de Cork - Teagasc
		CM4	? Laits fermentés et fromage	Calpis - Université de Kitasato
		reuteri	JCM 1112	Tractus digestif humain
	<i>rhamnosus</i>	100-23	Tractus digestif de rongeurs	JGI
ATCC 55730		Humain	BioGaia/Swedish Univ. of Agr. Science	
HN001		? Fromage	New Zealand Dairy Board	
<i>Pediococcus</i>	<i>pentosaceus</i>	ATCC 25745	Plantes	JGI / LABGC

La partie haute du tableau reprend la liste des BL pour lesquelles la séquence du génome est disponible. La partie basse correspond aux BL dont les génomes sont en cours de séquençage ou de publication.

? : Provenance de la souche inconnue. Dans ce cas, la niche écologique de l'espèce est précisée ; JGI : Joint Genome Institute ; LABGC : Lactic Acid Bacteria Genome Consortium ; KRIBB : Korean Research Institute of Bioscience and Biotechnology.

Adapté de Liu *et al.*, 2005.

Le troisième groupe est formé des genres *Oenococcus*, *Leuconostoc* et *Weissella*. Dans ce groupe, on retrouve des bactéries très utilisées dans l'industrie agroalimentaire comme *Oenococcus oeni*, bactérie importante dans la production de vin pour la fermentation malolactique ou *Leuconostoc citreum* utilisé pour la fermentation de certains légumes et *Lc. mesenteroides* ssp. *mesenteroides* utilisé dans la production de choucroute et de dextran.

Enfin, le dernier groupe phylogénétique est constitué des genres *Lactobacillus* et *Pediococcus*.

1.3. Les lactobacilles : un genre représentatif de la diversité des BL

Les lactobacilles représentent un genre important des BL tant au niveau industriel qu'au niveau de la flore commensale infantile. Cependant, ce genre est très hétérogène, aussi bien sur le plan génétique que sur le plan de la distribution des habitats.

1.3.1. Diversité génétique

Le genre *Lactobacillus* a été proposé par Beijerinck en 1901. Les critères phénotypiques restent encore très importants aujourd'hui pour la classification des lactobacilles (Schleifer et Ludwig, 1995). Le genre est composé de plus de 60 espèces décrites et est génétiquement assez divers. Le pourcentage en G+C varie entre 32 et 54 %. Une première classification des lactobacilles a été établie selon le profil fermentaire (de Roissart et Luquet, 1994, chap I-2) : homofermentaire strict (groupe I), hétérofermentaire facultatif (groupe II) et hétérofermentaire strict (groupe III). Cependant, cette classification ne permet pas de rendre compte des relations phylogénétiques entre espèces (Schleifer et Ludwig, 1995). Les études basées sur les ARN 16S ont permis de mettre en évidence l'existence de différents groupes phylogénétiques au sein des lactobacilles qui ne se recoupent pas systématiquement avec la classification basée sur les profils fermentaires (Schleifer et Ludwig, 1995). Les lactobacilles sont divisés en groupes phylogénétiques portant le nom de l'espèce type, considérée comme la plus représentative du groupe considéré : *L. acidophilus*, *L. salivarius*, *L. reuteri*, *L. buchneri*, *L. plantarum*... (Figure 2). *L. bulgaricus* appartient au groupe *L. acidophilus* (cf §1.3.4). Le genre *Pediococcus* ne peut pas être différencié du genre *Lactobacillus* par ce type de classification (Collins *et al.*, 1991), et est donc compris dans le groupe *Lactobacillus*.

Le premier génome de lactobacille à être publié a été celui de *L. plantarum* en 2003 (Kleerebezem *et al.*, 2003). *L. plantarum* est un lactobacille hétérofermentaire facultatif dont certaines souches sont utilisées comme probiotique (299v par exemple) ou en industrie agro-alimentaire (pour la production de légumes fermentés, levain, fromage...). Récemment, le génome de *L. sakei*, un autre lactobacille hétérofermentaire facultatif, phylogénétiquement éloigné de *L. plantarum*, a été séquencé (Chaillou *et al.*, 2005). Deux génomes de lactobacilles du groupe *acidophilus* (auquel appartient *L. bulgaricus*), *L. acidophilus* et *L. johnsonii* sont séquencés (Altermann *et al.*, 2005 ; Pridmore *et al.*, 2004). De nombreux projets de séquençage

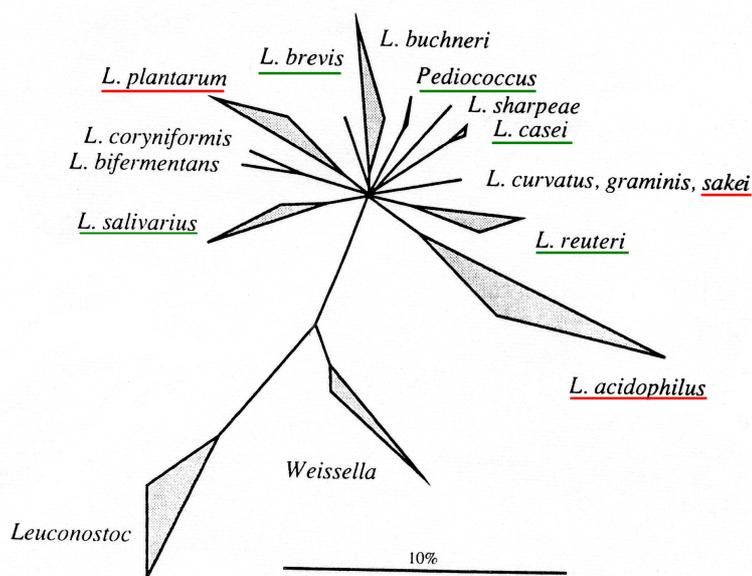


Figure 2 : Arbre phylogénétique des groupes principaux du genre *Lactobacillus* basé sur les ARNr 16S

Sont soulignés en rouge les groupes pour lesquels la séquence d'au moins un génome est disponible, en vert ceux pour lesquels au moins un génome est en cours de séquençage.

La barre indique une divergence de séquence estimée de 10%.

Adapté de Schleifer et Ludwig, 1995

sont en cours sur différents lactobacilles (Tableau 2 ; Liu *et al.*, 2005). Parmi les lactobacilles, la taille du génome varie entre 3,31 Mb pour *L. plantarum* à 1,86 Mb pour *L. bulgaricus*. Cependant, la majorité des génomes séquencés ou en cours de séquençage ont une taille d'environ 2 Mb (Liu *et al.*, 2005).

1.3.2. Diversité des habitats

Les lactobacilles ont colonisé un nombre important de niches écologiques (tableau 3) et représentent le groupe de BL le plus ubiquitaire dans l'environnement (de Roissart et Luquet, 1994, chap I-1). Certains lactobacilles peuvent être isolés de plusieurs niches écologiques. *L. plantarum* par exemple, se retrouve dans un grand nombre de produits fermentés (lait, viandes, légumes) ainsi que dans le tractus digestif humain et dans la salive. Certaines espèces sont limitées à des niches écologiques plus restreintes.

La grande diversité d'habitats explique que les lactobacilles sont très utilisés pour la production de nombreux produits fermentés traditionnels. Ils sont notamment utilisés pour les fermentations lactiques du lait (*L. bulgaricus*), de la viande (*L. sakei*) ou de produits végétaux (*L.*

plantarum). Les premières fermentations ont certainement été obtenues spontanément par contamination de l'aliment par les bactéries de l'environnement ou portées par l'aliment lui-même.

Tableau 3 : Habitats des lactobacilles

Environnement	Exemples
Plantes et produits végétaux	<i>L. plantarum</i> , <i>L. brevis</i> , <i>L. buchneri</i> , <i>L. fructosus</i>
Sols, eau, eaux usées et compost	<i>L. sharpeae</i> , <i>L. agilis</i> , <i>L. pentosus</i> , <i>L. rhamnosus</i>
Aliments fermentés	
lait	<i>L. delbrueckii</i> , <i>L. casei</i> , <i>L. helveticus</i> , <i>L. hilgardii</i> , <i>L. kefir</i>
viande	<i>L. sakei</i> , <i>L. curvatus</i> , <i>L. halotolerans</i> , <i>L. reuteri</i>
légumes	<i>L. plantarum</i> , <i>L. brevis</i> , <i>L. fermentum</i> , <i>L. cellobiosus</i>
Produits céréaliers	
levain (pain)	<i>L. delbrueckii</i> , <i>L. buchneri</i> , <i>L. acidophilus</i> , <i>L. casei</i> , <i>L. plantarum</i>
malt	<i>L. delbrueckii</i> , <i>L. casei</i> , <i>L. brevis</i>
Ensilage	<i>L. paracasei</i> , <i>L. plantarum</i> , <i>L. brevis</i> , <i>L. fermentum</i>
Humains	
Cavité orale	<i>L. plantarum</i> , <i>L. salivarius</i> , <i>L. oris</i> , <i>L. gasseri</i> , <i>L. crispatus</i>
Tractus digestif	<i>L. reuteri</i> , <i>L. gasseri</i> , <i>L. plantarum</i> , <i>L. ruminis</i> , <i>L. acidophilus</i>
Vagin	<i>L. crispatus</i> , <i>L. jensenii</i> , <i>L. gasseri</i> , <i>L. vaginalis</i> , <i>L. acidophilus</i>

Adapté de Stiles et Holzapfel, 1997 ; de Roissart et Luquet, 1994

Cependant, les lactobacilles peuvent aussi être des contaminants indésirables de produits alimentaires. Ils sont notamment à l'origine de la détérioration de produits tels que la bière (*L. casei*, *L. plantarum*), le lait (*L. maltaromicus*), la viande et les produits carnés (*L. sakei*, *L. curvatus*) (Stiles et Holzapfel, 1997 ; de Roissart et Luquet, 1994, chap VIII).

Certaines espèces de lactobacilles sont des bactéries commensales de l'Homme. Elles se situent au niveau de la cavité buccale, du tractus digestif et du tractus uro-génital, notamment dans le vagin. Ces bactéries sont importantes pour le maintien de l'équilibre de la flore commensale et donc pour la santé de l'hôte.

1.3.3. Effets des lactobacilles sur la santé

Depuis Metchnikoff et ses théories sur l'effet bénéfique des ferments lactiques, l'effet probiotique de certains lactobacilles a été et est encore largement étudié. Les probiotiques sont définis par la FAO (Food and Agriculture Organization of the United Nations) comme des micro-organismes vivants qui, lorsqu'ils sont administrés en quantité adéquate, confèrent un bénéfice pour la santé, au-delà de l'effet nutritionnel. Au sein d'une même espèce, les propriétés probiotiques de souches différentes peuvent être très variables (Isolauri *et al.*, 2004).

Les lactobacilles font partie de la flore digestive et participent aux effets bénéfiques de la microflore sur l'hôte, que ce soit au niveau métabolique, trophique ou immunitaire (décrits § 1.1.4). Un des effets majeurs des lactobacilles sur la santé humaine décrits dans la littérature est leur rôle antagoniste vis à vis des bactéries pathogènes (Servin, 2004). Ce rôle a été démontré dans le cas de pathogènes du tractus digestif et du tractus uro-vaginal. Les mécanismes impliqués dans cet antagonisme ont été mis en évidence pour certaines souches de lactobacilles commensaux ou ayant des propriétés probiotiques. Les effets des lactobacilles sur les bactéries pathogènes reposent sur différentes propriétés telles que la capacité à adhérer aux cellules et d'inhiber ainsi l'adhésion des pathogènes, l'inhibition de la croissance des pathogènes, la capacité d'entrer en compétition avec les pathogènes pour les ressources nutritives et la modulation de la réponse immunitaire de l'hôte (Reid et Burton, 2002). De très nombreuses études ont été menées et ont décrit ces différentes propriétés chez différentes souches de lactobacilles. Je me focaliserai ici sur les mécanismes et facteurs connus impliqués dans l'antagonisme des lactobacilles vis à vis des bactéries pathogènes.

Certains lactobacilles sont capables d'induire la réponse immunitaire de l'hôte et de la moduler. Cet effet a été décrit pour plusieurs souches de lactobacilles comme *L. johnsonii* La1, *L. casei rhamnosus* GG ou *L. casei* Shirota. Cependant, les fonctions responsables de l'immunomodulation par les bactéries commensales ou probiotiques sont largement inconnues. Parmi les facteurs pouvant jouer un rôle dans l'immunomodulation, celui de la composition des acides lipotéchoïques a été démontré (Grangette *et al.*, 2005).

L'adhésion des lactobacilles aux cellules humaines serait importante pour deux raisons principales : elle permettrait aux lactobacilles de se maintenir dans l'hôte et, *in vitro*, d'inhiber l'adhésion des pathogènes. De façon générale, le processus d'adhésion fait intervenir des interactions électrostatiques, des forces stériques et de nature hydrophobe ainsi que les acides lipotéchoïques et des structures spécifiques comme la lectine ou des structures protéiques extracellulaires (couche S) (Servin, 2004). Peu de facteurs d'adhésion aux cellules humaines sont connus chez les lactobacilles. Chez *L. fermentum* 104R, une protéine de surface de 29kDa qui se lie à la mucine gastrique et au mucus intestinal de porc a été caractérisée (Rojas *et al.*, 2002). Une séquence partielle de cette protéine a été déterminée mais elle ne correspond à aucune protéine connue. Chez *L. johnsonii* La1, l'acide lipotéchoïque semble impliqué dans l'adhésion aux cellules de type enterocyte Caco-2 (Granato *et al.*, 1999). Récemment, une étude par

mutation dirigée de gènes candidats chez *L. acidophilus* NCFM a montré que chez cette bactérie, plusieurs déterminants sont impliqués dans l'adhésion de cette bactérie à des cellules de type Caco-2. L'inactivation de trois gènes diminue fortement l'adhésion, *mub* codant une protéine de liaison à la mucine, *fbpA* codant une protéine de liaison à la fibronectine et *slpA* codant une protéine de couche S (S-layer) mais, pour cette dernière protéine, l'effet serait indirect (Buck *et al.*, 2005).

La capacité des lactobacilles à inhiber la croissance de bactéries pathogènes repose sur plusieurs propriétés. La production de peroxyde d'hydrogène (H₂O₂) comme facteur inhibiteur de la croissance des pathogènes a surtout été étudiée pour les lactobacilles commensaux du tractus urogénital. Elle semble être un mécanisme de défense aspécifique employé par la microflore vaginale (Reid et Burton, 2002). La microflore vaginale des femmes souffrant d'infections vaginales bactériennes est constituée en majorité de lactobacilles commensaux ne produisant pas d'H₂O₂ (seuls 14% des lactobacilles en produisent) alors que pour les femmes « saines », les lactobacilles producteurs d'H₂O₂ représentent 75% de la flore (Al-Mushrif et Jones, 1998). Il a été montré que *L. jensenii* et *L. crispatus*, deux lactobacilles dominants de la flore vaginale, inhibent la croissance de *Neisseria gonorrhoeae* par production d'H₂O₂ (expériences *in vitro* ; St Amant *et al.*, 2002).

Un autre facteur permettant l'inhibition de la croissance des pathogènes est la production d'acides organiques (lactique ou acétique), et par là-même l'acidification du milieu. Ces effets ont été étudiés *in vitro* pour plusieurs lactobacilles probiotiques. *L. casei rhamnosus* GG inhibe l'adhésion de *Salmonella typhimurium* par un effet dépendant du pH, l'inhibition étant levée à pH neutre (Lehto et Salminen, 1997). *L. casei* Shirota et *L. acidophilus* YIT0070 inhibent la croissance de *E. coli* O157:H7 et cette inhibition dépend de la production d'acide lactique (Ogawa *et al.*, 2001). Cependant, il est difficile de différencier les effets du pH de ceux de l'acide lactique. Fayol-Messaoudi et coll. (2005) ont montré que l'activité antagoniste de *L. johnsonii* La1, *L. casei rhamnosus* GG, *L. rhamnosus* GR1, *L. casei* Shirota et *L. casei* DN-114 001 sur *Salmonella enterica* Typhimurium est principalement due à l'effet pH. Cependant, il semble que l'acide lactique joue un rôle dans l'inhibition de croissance des bactéries à Gram négatif (auxquelles appartiennent un grand nombre de pathogènes) par perméabilisation de la membrane externe, ce qui faciliterait l'entrée de molécules antibactériennes dans les cellules et augmenterait ainsi leur efficacité (Alakomi *et al.*, 2000). De plus, l'accumulation de lactate dans la cellule induit aussi des effets délétères pour la cellule (Presser *et al.*, 1997).

Certains lactobacilles produisent des molécules protéiques bactéricides appelées bactériocines dont le rôle dans l'inhibition de pathogènes en conditions technologiques a été largement étudié. En 1993, Klaenhammer définit trois groupes de bactériocines (Klaenhammer, 1993). Les bactériocines de la classe I sont de petites bactériocines composées d'un ou deux peptides d'environ 3 kDa. Ces peptides sont soumis à des modifications post-traductionnelles de sorte à contenir de la lanthionine, de la β -methyllanthionine et des acides aminés déshydratés (Ross *et al.*, 2002). Les bactériocines de cette classe sont aussi appelés lantibiotiques. La nisine appartient à cette famille. La classe II correspond à de petits peptides (<5kDa) non modifiés, thermostables et qui sont divisés en trois sous-groupes, (1) un groupe actif contre *Listeria*, (2) les peptides nécessitant deux composants pour leur activité, et (3) les peptides sécrétés par le système de sécrétion *sec* (les autres sont sécrétés par un transporteur spécifique). La sakacine A appartient à cette famille. La classe III est constituée par des molécules de plus haut poids moléculaire (>30kDa) qui sont thermolabiles (Servin, 2004). Les bactériocines ont un spectre d'action variable, le plus souvent limité à certaines espèces à Gram (+). Certaines bactériocines peuvent éventuellement être utiles pour limiter les contaminations alimentaires par des pathogènes comme *Listeria monocytogenes* ou *Clostridium botulinum* (Ross *et al.*, 2002), bien que leur activité soit souvent limitée à une gamme de pH donné et que des bactéries résistantes puissent apparaître. L'usage des bactériocines pour le traitement de troubles intestinaux est limité car elles sont peu ou pas actives sur des bactéries à Gram négatif, principaux agents de ces troubles. Cependant, quelques bactériocines de lactobacilles sont actives sur des bactéries à Gram négatif comme l'acidophiline 801 (Zamfir *et al.*, 1999), active contre les pathogènes *Escherichia coli* Raw et *Salmonella panama*.

Différentes espèces de lactobacilles sont largement étudiées pour leurs effets bénéfiques sur la santé. Cependant, certaines de ces bactéries ont été impliquées dans plusieurs cas de bactériémies, d'endocardites ou d'infections localisées (Cannon *et al.*, 2005). Il a été montré que dans certains cas, les souches responsables de bactériémies ont des profils en PFGE identiques à une souche utilisée comme probiotique, *L. rhamnosus* GG (Salminen *et al.*, 2004). Les infections dues aux lactobacilles sont rares mais, comme tout micro-organisme, les lactobacilles peuvent être associés à des infections, dans certaines conditions, notamment dans le cas de sujets immunodéprimés ou déjà malades (Salminen et Arvilommi, 2002). Aujourd'hui, le débat reste encore ouvert sur l'innocuité des probiotiques, qui pourraient être des opportunistes notamment chez les patients immuno-déprimés (Cannon *et al.*, 2005).

1.3.4. Le groupe *acidophilus*

Parmi les différents groupes constituant le genre *Lactobacillus* (figure 2), le groupe *acidophilus* nous intéresse particulièrement puisqu'il contient l'espèce *L. bulgaricus*. Le groupe *acidophilus* comprend en majorité des homofermentaires stricts cependant certains d'entre eux sont des hétérofermentaires facultatifs comme *L. acetotolerans* ou *L. intestinalis* (Stiles et Holzapfel, 1995). Le pourcentage en G+C de ce groupe est relativement variable, entre 33 et 51% (figure 3). *L. delbrueckii* est atypique car particulièrement riche en G+C (49-51%), par rapport aux autres espèces du groupe. Ce groupe s'appelait le groupe *L. delbrueckii* (Collins *et al.*, 1991) mais a été renommé *L. acidophilus*, cette espèce étant plus représentative du groupe en ce qui concerne le %G+C (Stiles et Holzapfel, 1995).

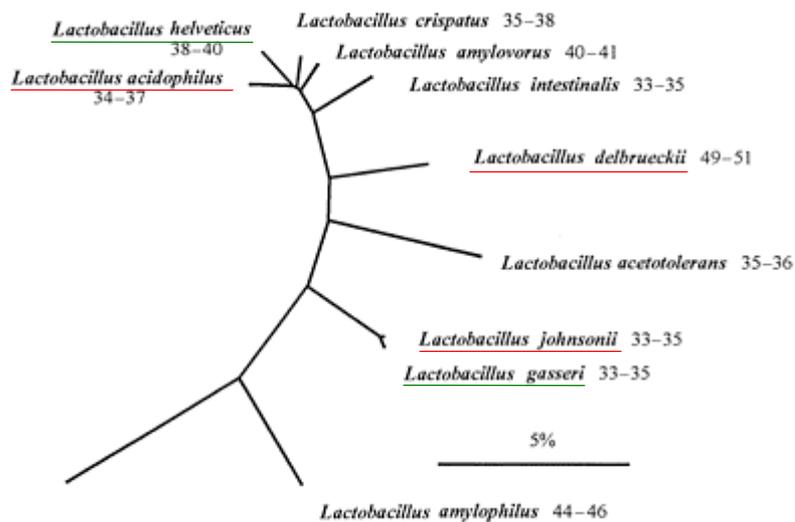


Figure 3 : Arbre phylogénétique du groupe *L. acidophilus*.

Les nombres situés à côté des noms des espèces représentent le %GC de chaque espèce. La barre indique une divergence de séquence estimée de 5%.

Sont soulignés en rouge les espèces pour lesquelles la séquence d'au moins un génome est disponible, en vert celles pour lesquelles au moins un génome est en cours de séquençage.

Adapté de Schleifer et Ludwig, 1995

Ce groupe contient aussi bien des bactéries très utilisées industriellement dans la transformation du lait (*L. delbrueckii* ssp. *bulgaricus* et *lactis*, *L. helveticus*), des bactéries issues de végétaux (*L. amylovorus*, *L. delbrueckii* ssp. *delbrueckii* et *L. amylophilus*) que des bactéries commensales de l'homme (*L. acidophilus*, *L. crispatum*, *L. johnsonii* et *L. gasseri*). Certaines souches de ces lactobacilles commensaux sont étudiées pour leurs propriétés probiotiques. Des études cliniques semblent montrer que la consommation de *L. acidophilus* NAS, *L. johnsonii*

La1 et *L. gasseri* OLL 2716 aurait un effet positif contre les infections à *Helicobacter pylori*. Utilisés comme supplément aux traitements thérapeutiques classiques des infections à *H. pylori*, *L. johnsonii* La1 et *L. acidophilus* LB pourraient améliorer le traitement ou diminuer l'activité du pathogène (Reid *et al.*, 2003 ; Hamilton-Miller, 2003). *L. crispatus* CTV-05 est en phase d'essais cliniques pour la prévention d'infections vaginales (Hoesl et Altwein, 2005 ; Reid *et al.*, 2003).

L'intérêt porté à ce groupe de lactobacilles est visible par l'effort de séquençage des génomes de différentes souches appartenant à ce groupe, puisque les génomes d'au moins sept souches appartenant à cinq espèces différentes sont séquencés ou en cours de séquençage (Tableau 2, Figure 3).

Dans cette première partie, nous avons vu qu'une grande partie des connaissances sur les lactobacilles est basée sur des données physiologiques et descriptives. Récemment, les séquences de plusieurs lactobacilles ont été publiées et ont permis d'apporter un éclairage nouveau sur ce genre.

2. Apports de la génomique sur l'étude des bactéries lactiques

La génomique est l'étude de l'ensemble de l'information génétique d'un organisme. La génomique est une discipline récente et en pleine expansion. Le premier génome bactérien séquencé et publié est celui de *Haemophilus influenzae*, agent pathogène responsable de méningites et dont le génome fait 1,8 Mb (Fleischmann *et al.*, 1995). La même année, le génome de *Mycoplasma genitalium*, agent d'infections uro-génitales qui possède le plus petit génome bactérien connu (580 kb) est publié (Fraser *et al.*, 1995). Cette publication donnera déjà lieu à la comparaison des deux seuls génomes disponibles, et ouvrira la voie de la génomique comparative. De 1995 à 2000, 32 génomes bactériens sont publiés dont ceux d'*Escherichia coli* (Blattner *et al.*, 1997), bactérie modèle des Gram (-) et de *Bacillus subtilis* (Kunst *et al.*, 1997), bactérie modèle des Gram (+). Les progrès techniques et la création de grands centres dédiés au séquençage ont permis à la génomique de se développer. De 2001 à 2005, 237 génomes dont 63 pour l'année 2005 ont été séquencés et publiés (source : <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>).

2.1. Génomique et génomique comparative des capacités d'adaptation des souches

2.1.1. Informations issues d'un génome

2.1.1.1. Structure du génome

La connaissance de la séquence d'un génome nous donne accès à un grand nombre d'informations parmi lesquelles la structure du génome étudié. La connaissance de la structure des génomes bactériens a largement profité des progrès de la génomique. Les paramètres les plus utilisés pour la description de la structure des génomes sont la taille, la géométrie, le nombre de réplicons et le pourcentage en G+C. La taille des génomes bactériens séquencés varie de 580 kb (*Mycoplasma genitalium*) à 9,1 Mb (*Bradyrhizobium japonicum*) (Bentley et Parkhill, 2004). La taille des génomes est la résultante de différents évènements comme la duplication de gènes, l'acquisition de matériel génétique par transfert horizontal et la perte de gènes par délétion (Mira

et al., 2001). Les plus petits génomes appartiennent à des espèces vivant dans des niches écologiques stables, le plus souvent en association avec un hôte. Les plus grands génomes, à l'inverse, se retrouvent dans des espèces vivant dans des milieux plus complexes et variables, comme le sol par exemple. La densité des gènes varie peu (environ 1 gène par kb), la taille du génome est donc en général proportionnelle au nombre de gènes présents (Bentley et Parkhill, 2004).

La majorité des espèces bactériennes possède un chromosome circulaire unique. Cependant, certaines espèces possèdent un chromosome linéaire comme celles des genres *Streptomyces* (Kieser *et al.*, 1992) ou *Borrelia* (Ferdows et Barbour, 1989). Certaines espèces possèdent plusieurs chromosomes, codant chacun des fonctions essentielles à la survie cellulaire, par exemple les opérons ribosomiques (Ochman, 2002). *Agrobacterium tumefaciens* contient des chromosomes linéaires et circulaires (Allardet-Servent *et al.*, 1993).

La génomique a révélé que la composition nucléotidique des génomes n'est pas homogène, il est proposé que cela résulte de contraintes qui sont à ce jour encore mal comprises.

Le pourcentage en G+C des génomes séquencés se situe entre 26,5% pour *Wigglesworthia glossinidia* (Akman *et al.*, 2002) et 72,1% pour *Streptomyces coelicolor* (Bentley *et al.*, 2002). Une corrélation semble exister entre la taille des génomes et le pourcentage en G+C mais son explication n'est pas claire aujourd'hui (Bentley et Parkhill, 2004). Au sein d'un même chromosome, le pourcentage en G+C varie. Les séquences intergéniques sont plus riches en A+T, et notamment les régions promotrices des gènes, facilitant ainsi l'ouverture des brins d'ADN lors de la transcription (Pedersen *et al.*, 2000). Certainement pour des raisons similaires, l'origine de réplication est, elle aussi, riche en A+T (Bentley et Parkhill, 2004). Il existe aussi des régions plus étendues pour lesquelles le pourcentage en A+T est plus élevé que la moyenne du génome. Ces régions de composition atypique sont souvent assimilées à de l'ADN issu de transfert horizontal (Daubin *et al.*, 2003). Un autre paramètre est souvent regardé dans les génomes, le «GC skew». Le brin précoce est plus riche en G qu'en C, et inversement pour le brin retardé. Cela crée une inversion du «GC skew» (G-C/G+C) à l'origine de réplication et au terminus, ce qui permet de les situer sur le chromosome (McLeane *et al.*, 1998).

L'obtention de la séquence génomique d'un organisme permet aussi de mettre en évidence des éléments pouvant intervenir dans la plasticité du génome. Les séquences répétées (rDNA, séquences d'insertion) sont susceptibles de recombiner entre elles et d'induire des réarrangements dans le génome (Hughes, 2000). Les éléments mobiles sont une autre source de plasticité du génome. Parmi ces éléments, on trouve les prophages, les transposons, les

séquences d'insertion et les éléments conjugatifs qui lors de leur excision du génome peuvent emmener de l'ADN de l'hôte qui sera intégré ensuite dans le chromosome receveur (Dobrindt *et al.*, 2004). La compétence naturelle est aussi un élément qui intervient dans la plasticité des génomes puisqu'elle permet à la bactérie compétente d'introduire dans son génome de l'ADN issu du milieu extérieur.

2.1.1.2. Capacités métaboliques

La séquence d'un génome donne accès à l'ensemble des gènes présents dans un organisme. Cependant, l'identification d'un cadre ouvert de lecture ne permet pas toujours d'avoir accès à sa fonction. Encore aujourd'hui, 20 à 40% des gènes détectés dans les génomes sont de fonction inconnue (Zheng *et al.*, 2002). La génomique est donc tributaire des avancées dans les autres disciplines comme la physiologie ou la génétique classique. Lorsqu'un gène est détecté dans un génome, la première étape est d'attribuer une fonction (putative) au produit du gène. La plupart du temps, cette fonction est attribuée par homologie avec des gènes déjà étudiés, par identification de domaines conservés avec d'autres protéines ou plus occasionnellement par rapport à un contexte génétique (Gerlt et Babbitt, 2000). Ce processus, appelé annotation, est une partie essentielle à l'utilisation des données génomiques. L'annotation présente plusieurs limites : seules les fonctions déjà connues peuvent être identifiées, la fonctionnalité de la protéine n'est que putative (possibilité de mutations qui inactivent la protéine ce qui génère des faux positifs), et dans le cas de remplacements non orthologues (fonction assurée par une protéine non homologue), une fonction peut ne pas être identifiée (génère des faux négatifs).

Sur la base des fonctions putatives assignées aux gènes, il est possible de reconstruire *in silico* les capacités métaboliques de l'organisme. Certaines bases de données se sont spécialisées dans cette démarche, comme KEGG (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.ad.jp/kegg/>), ERGO (<http://ergo.integratedgenomics.com/ERGO/>) ou MetaCyc (<http://metacyc.org/>). La reconstruction des voies métaboliques présentes dans un génome a été réalisée pour plusieurs organismes (Borodina et Nielsen, 2005). Pour *L. plantarum*, la reconstruction des voies métaboliques a été effectuée dans l'environnement Pathway Tools (Karp *et al.*, 2002) à partir des EC numbers (Enzyme Classification number : numéro permettant de caractériser la réaction réalisée par une enzyme) issus de l'annotation et des noms des produits des gènes (Teusink *et al.*, 2005). Cette reconstruction automatique a ensuite été vérifiée manuellement et trois grands types de modifications ont été apportés : (i) identifier et résoudre les incohérences (identifier des candidats pour les fonctions manquantes dans les voies de biosynthèse), (ii) définir la notion de complexe protéique nécessaire à une réaction et vérifier que

toutes les sous-unités sont présentes et (iii) ajouter les voies métaboliques spécifiques de *L. plantarum*. Le résultat des reconstructions des voies de biosynthèse des acides aminés et des vitamines a ensuite été comparé avec les exigences nutritionnelles de la souche (Teusink *et al.*, 2005). Après vérification manuelle des reconstructions et modifications, seules cinq prédictions sur 37 se sont révélées en contradiction, dans les conditions testées, avec les expériences d'exigences nutritionnelles (contre 13/37 pour les prédictions automatiques non validées manuellement). Les reconstructions métaboliques représentent donc un moyen puissant de prédiction de voies métaboliques fonctionnelles chez un organisme à partir des données génomiques. Cependant, ces reconstructions, pour être fiables, nécessitent un travail de nettoyage manuel important et ne peuvent être effectuées que sur des voies de biosynthèse bien caractérisées.

L'étape supplémentaire dans l'analyse des données génomiques est la comparaison entre plusieurs génomes.

2.1.2. Comparaison de génomes

La comparaison de génomes de souches différentes de la même espèce permet de mettre en évidence les régions variables entre les souches. Cette comparaison a surtout été réalisée pour des génomes de bactéries pathogènes. Chez ces bactéries, il était connu de longue date que les différences de phénotype entre deux souches étaient souvent dues à des gènes groupés physiquement dans le génome (notion de « cluster »). Ces régions génomiques, appelées « îlots de pathogénicité » (Hacker *et al.*, 1990) ou « îlots génomiques » (Boyd *et al.*, 2000b) suivant qu'ils contiennent ou non des gènes associés à la virulence, ont été retrouvées dans les comparaisons entre génomes de souches d'une même espèce, comme dans le cas de *E. coli* (pour revue : Whittam et Bumbaugh, 2002).

Chez les bactéries lactiques, la seule comparaison de génomes de deux souches d'une même espèce à être publiée, a été réalisée pour *Streptococcus thermophilus* (Bolotin *et al.*, 2004). Les deux souches, isolées de yaourt, ont plus de 90% des séquences codantes en commun et diffèrent par 55 insertions ou délétions de plus de 50 bases qui représentent 4% de la longueur du génome. 23 différences sont liées à des séquences d'insertion (IS) qui sont présentes à des endroits différents entre les deux génomes. Ces insertions sont soit restreintes à l'IS (pour 18 d'entre elles), soit accompagnées de groupes de gènes supplémentaires (pour 5 d'entre elles). Les gènes qu'on retrouve associés aux IS sont les groupes de gènes de biosynthèse des exopolysaccharides (EPS) dans deux cas, des groupes de gènes de fonction inconnue dans deux cas et un groupe de

gènes impliqués dans la biosynthèse de bactériocine. Les autres différences les plus importantes non associées à des éléments génétiques mobiles concernent des gènes impliqués dans la biosynthèse des polysaccharides à rhamnose (4 opérons), un prophage, l'immunité et la synthèse de bactériocines (4 loci).

Une étude portant sur la diversité génomique entre 20 souches de *L. plantarum* a été réalisée par puces à ADN (Molenaar *et al.*, 2005). Cette comparaison est partielle puisque l'ADN présent sur les puces représente 80% du génome de la souche *L. plantarum* WCFS1 séquencée (Kleerebezem *et al.*, 2003). Les séquences manquantes dans les 19 autres souches peuvent donc être détectées mais pas les séquences supplémentaires. Parmi les régions qui sont absentes d'au moins 30% des souches, on retrouve en particulier des clusters de gènes dont les produits sont impliqués dans l'utilisation de sucres (9 clusters), la synthèse de protéines extracellulaires (3 clusters), la synthèse non ribosomale de protéines (1 cluster) et la synthèse d'un cofacteur, la molybdoptérine (1 cluster). On retrouve aussi des groupes de gènes classiquement variables entre souches comme les deux opérons de synthèse d'EPS, trois prophages, un cluster de production d'une bactériocine, la plantaricine et un système de restriction-modification. Aucune corrélation n'a pu être mise en évidence entre la variabilité génomique et la provenance des souches (10 provenant d'isolats humains et 9 de produits végétaux fermentés) par cette méthode.

La comparaison de génomes peut aussi être réalisée entre les génomes d'espèces différentes mais proches phylogénétiquement. Chez les lactobacilles, jusqu'à aujourd'hui, les séquences de seulement quatre génomes sont disponibles, tous d'espèces différentes. Les espèces les plus proches pour lesquelles les génomes ont été séquencés sont *L. johnsonii* (Pridmore *et al.*, 2004) et *L. acidophilus* (Altermann *et al.*, 2005). Aucune étude n'a été publiée sur la comparaison de ces deux génomes. Cependant certaines caractéristiques sont communes aux deux génomes séquencés. La taille des deux génomes est d'environ 2 Mb et le pourcentage en G+C est de 34,71% pour *L. acidophilus* NCFM et 34,6% pour *L. johnsonii* NCC 533. Les deux bactéries semblent incapables de synthétiser la plupart des acides aminés et des vitamines mais possèdent des protéases et un grand nombre de transporteurs de peptides et d'acides aminés et de peptidases. Les deux génomes codent pour de nombreux systèmes de transports et de métabolisme de sucres, notamment d'hexoses. *L. johnsonii* NCC 533 a été isolée de l'intestin humain (Haller *et al.*, 2002) et *L. acidophilus* NCFM a été isolée de fèces d'enfant (Altermann *et al.*, 2005) et sont toutes les deux étudiées pour leur propriétés probiotiques. Leurs caractéristiques génomiques communes semblent liées à l'environnement de ces deux souches, le

tractus digestif, qui est un milieu riche où les bactéries peuvent se procurer un grand nombre de nutriments.

La comparaison entre les génomes de *L. plantarum* et *L. johnsonii* a été publiée en 2004 (Boekhorst *et al.*, 2004). Ces deux espèces sont plus éloignées phylogénétiquement que *L. johnsonii* et *L. acidophilus* (figure 2). La souche *L. plantarum* WCFS1 provient aussi d'un isolat humain, puisqu'elle a été isolée de la salive. Les deux génomes ont des caractéristiques générales très différentes. La souche de *L. plantarum* possède un génome de plus de 3,3 Mb par rapport à 2 Mb pour *L. johnsonii*. Le pourcentage en G+C est lui aussi très différent, de 45,6% à 34,6% respectivement. La synténie entre les génomes (i.e. l'ordre des gènes dans le génome) n'est pas globalement conservée, seuls certains groupes de gènes restent associés dans les deux génomes. Au niveau des capacités métaboliques, alors que *L. johnsonii* est incapable de synthétiser la plupart des acides aminés, *L. plantarum* possède les voies de biosynthèse de tous les acides aminés hormis la leucine, la valine et l'isoleucine. De la même manière, *L. plantarum* est capable de synthétiser certaines vitamines comme le folate et peut métaboliser une plus large variété de sucres que *L. johnsonii*. Toutes ces observations ont été rapprochées des niches écologiques dans lesquelles évoluent les deux espèces. *L. johnsonii* est une bactérie que l'on retrouve quasi exclusivement dans le tractus digestif alors que *L. plantarum* est capable de survivre dans différents écosystèmes, des plantes au tractus digestif. Sa plus grande polyvalence se retrouve au niveau génomique, avec un métabolisme plus complexe et une plus grande autonomie vis à vis du milieu extérieur (moins d'auxotrophies).

La comparaison génomique entre des bactéries phylogénétiquement distantes mais partageant le même habitat peut améliorer la compréhension des mécanismes d'adaptation des bactéries à leur environnement. Peu d'études ont été menées sur le sujet jusqu'à présent. Plusieurs séquences de génomes de bactéries dont certaines souches peuvent coloniser le tractus digestif sont disponibles parmi lesquelles *L. plantarum* (Kleerebezem *et al.*, 2003), *Enterococcus faecalis* (Paulsen *et al.*, 2003), *Bifidobacterium longum* (Schell *et al.*, 2002), *L. monocytogenes* (Glaser *et al.*, 2001), *Bacteroides thetaiotamicron* (Xu *et al.*, 2003), *Escherichia coli* (Blattner *et al.*, 1997) et une première comparaison des caractéristiques de ces génomes a été réalisée par de Vos *et coll.* (de Vos *et al.*, 2004). Les bactéries du tube digestif possèdent un très grand nombre de systèmes de transport et de dégradation des sucres mais ceux-ci diffèrent selon la localisation des bactéries dans le tractus digestif (figure 4).

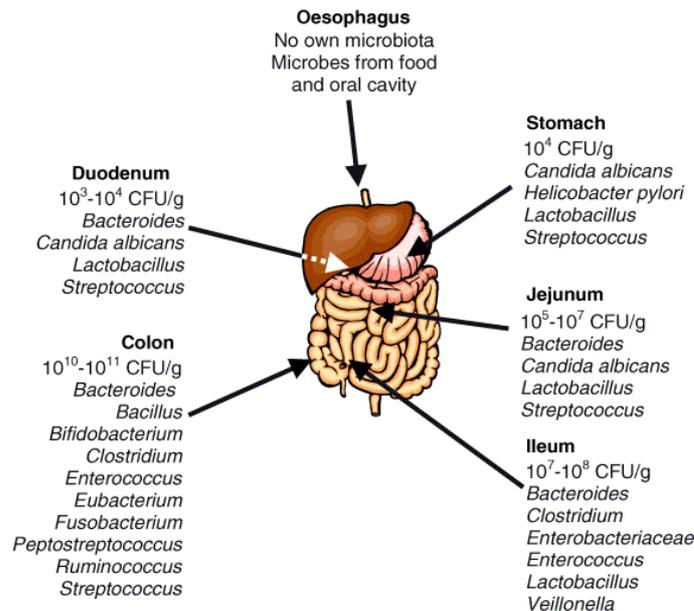


Figure 4: Les genres dominants dans les différents compartiments du tractus digestif (Isolauri *et al.*, 2004)

Les bactéries présentes dans le colon (*B. thetaiotamicron* et *B. longum*) possèdent des systèmes d'utilisation de sources de carbone complexes notamment des glycosyl hydrolases qui permettent la dégradation d'hydrates de carbone issus de végétaux ou produits par l'hôte (mucus par exemple). Ces caractéristiques pourraient être importantes pour la survie de ces bactéries dans le colon, milieu pauvre en mono- et disaccharides utilisés en amont par l'hôte et la microflore vivant dans la partie supérieure du tractus (duodenum, jejunum et iléum). Sur la base des génomes séquencés, une caractéristique commune aux bactéries du tube digestif semble être le nombre important de systèmes de régulation dédiés à l'utilisation des hydrates de carbone. Les systèmes de régulation mis en jeu varient entre les espèces (facteurs sigma alternatifs, systèmes à deux composants, activateurs ou répresseurs). De plus, ces bactéries interagiraient avec l'hôte via leur enveloppe cellulaire. Les génomes de *L. plantarum*, *B. longum* et *B. thetaiotamicron* codent pour un grand nombre de protéines extracellulaires parmi lesquelles se trouvent des protéines qui seraient impliquées dans la liaison au mucus ou à la fibronectine pour *L. plantarum* et des structures de type fimbriae de liaison aux glycoprotéines pour *B. longum*.

La comparaison de génomes, qu'elle soit réalisée entre des souches proches phylogénétiquement ou non, a permis de mettre en évidence une certaine plasticité des génomes et notamment les transferts horizontaux. Ces transferts de matériel génétique peuvent représenter une grande partie d'un génome bactérien, comme le montre la comparaison de trois souches

différentes d'*E. coli* qui n'ont qu'environ 40% de leurs gènes en commun (Welch *et al.*, 2002). Dans ce cas particulier, un grand nombre de différences entre les génomes est dû à des îlots génomiques ou de pathogénicité et à des prophages.

Deux méthodes sont utilisées pour détecter les gènes acquis par transfert horizontal : l'approche phylogénétique et l'approche paramétrique (Lawrence et Ochman, 2002).

L'approche phylogénétique se base sur la distribution ou la forte similarité d'un gène avec des homologues dans des bactéries non apparentées (Ragan, 2001 ; Lawrence et Hartl, 1992). Un gène acquis par transfert horizontal aurait donc une histoire phylogénétique différente de celle de son hôte. Cette méthode a besoin, pour être utilisée à l'échelle d'un génome, de plusieurs séquences génomiques de bactéries proches, voire de celles de plusieurs souches d'une même espèce.

Les méthodes paramétriques sont basées sur l'hypothèse que les gènes dont la composition nucléique et l'utilisation des codons synonymes sont différentes de celles de la majorité des gènes de l'organisme, sont probablement issus de transferts horizontaux (Mrazek et Karlin, 1999 ; Medigue *et al.*, 1991 ; Moszer *et al.*, 1999). Ces méthodes ne nécessitent pas d'autres données génomiques que celles de l'organisme. Cependant, elles sont incapables de détecter des gènes issus de transfert horizontaux anciens (les gènes sont soumis au même biais mutationnel une fois intégrés dans le génome receveur et deviennent de moins en moins atypiques) ainsi que les gènes provenant de bactéries ayant le même biais mutationnel et la même utilisation de codons synonymes. Cette méthode détecte donc de manière plus fiable les gènes acquis récemment et sous-estime probablement leur nombre.

Ces deux méthodes, utilisées pour le génome d'*E. coli* ne donnent pas les mêmes résultats (Koski *et al.*, 2001). Néanmoins, les résultats issus des deux analyses diffèrent plus sur la taille des régions transférées et donc sur le nombre de gènes que sur le nombre de transferts à proprement parler (Lawrence et Ochman, 2002).

2.1.3. *Génome minimal et ancêtre commun*

Une question est apparue avant le développement de la génomique : parmi les gènes détectés, combien sont indispensables à la vie cellulaire ? Le génome minimal peut être défini comme le plus petit ensemble de gènes suffisants pour permettre une forme de vie cellulaire fonctionnelle dans les meilleures conditions possibles, c'est-à-dire en présence de tous les nutriments essentiels et en absence de stress environnementaux (Gil *et al.*, 2004). Différentes approches ont été utilisées pour identifier les gènes essentiels à la survie cellulaire. Certaines

sont des approches expérimentales par inactivation de gènes, d'autres sont des méthodes bioinformatiques par comparaison de génomes. Une limite supérieure du nombre de gènes essentiels à la croissance cellulaire est donnée par le nombre de gènes présents dans le plus petit génome connu : 480 gènes pour *M. genitalium*, un parasite intracellulaire.

Dès 1996, une étude sur les deux génomes séquencés à cette date, *M. genitalium* (Fraser *et al.*, 1995) et *H. influenzae* (Fleischmann *et al.*, 1995) est effectuée pour déterminer le groupe de gènes constituant le génome minimal (Mushegian et Koonin, 1996). L'hypothèse de départ est que les gènes nécessaires à la survie cellulaire sont conservés entre espèces, et notamment entre les deux grands groupes de bactéries (Gram positif et négatif) auxquels appartiennent les deux espèces. Cette comparaison montre que seuls 240 gènes sont conservés entre les deux espèces. A ce groupe, 22 gènes ont été ajoutés pour prendre en compte des fonctions présentes dans les deux souches mais remplies par des protéines non homologues (comme la phosphoglycérate mutase ou la nucléoside diphosphate kinase). 6 gènes redondants ou spécifiques du style de vie des deux bactéries (comme les gènes du système PTS fructose ou l'hémolysine) ont été enlevés à ce groupe de gènes. Au final, Mushegian et Koonin proposent que le génome minimal contienne environ 256 gènes. Récemment, une étude de mutagenèse globale par transposition a permis d'identifier les gènes essentiels chez *M. genitalium* (Glass *et al.*, 2006). Sur 482 gènes, 387 semblent essentiels à la survie de la bactérie, ce qui est largement supérieur aux 256 gènes prédits par Mushegian et Koonin. Parmi les 387 gènes, 110 codent des protéines de fonctions inconnues.

En 2003, une étude par inactivation systématique des gènes de *Bacillus subtilis* (intégrants simples ne présentant qu'une mutation) montre que 271 gènes sont essentiels à la croissance de la bactérie (Kobayashi *et al.*, 2003). Cependant, cette étude présente deux limites : les fonctions codées de manière redondante ne sont pas détectées comme essentielles et certains gènes peuvent être essentiels chez *B. subtilis* sans faire partie du groupe minimal de gènes (par exemple, certaines enzymes de la voie respiratoire sont essentielles chez *B. subtilis* alors qu'elles sont absentes de nombreux génomes, dont ceux des lactobacilles du groupe *acidophilus* et seuls 57% des gènes essentiels de *B. subtilis* sont conservés chez les bactéries possédant des petits génomes comme les mycoplasmes). 146 (sur 271) gènes considérés comme essentiels chez *B. subtilis* sont conservés chez *M. genitalium* où 8 d'entre eux ont pu être inactivés (Glass *et al.*, 2006). En majorité, les groupes fonctionnels dans lesquels se retrouvent les gènes essentiels de *B. subtilis* coïncident avec ceux identifiés par Mushegian et Koonin (1996), c'est à dire le métabolisme de

l'ADN et de l'ARN, la synthèse protéique, la synthèse de l'enveloppe cellulaire, la glycolyse, le métabolisme des nucléotides et des coenzymes et cofacteurs.

Une autre étude, portant sur l'inactivation de gènes chez *Staphylococcus aureus* par l'utilisation d'ARN anti-sens, a montré que 658 gènes étaient essentiels à la croissance de la bactérie (Forsyth *et al.*, 2002). Sur ces 658 gènes, 168 ont un homologue chez *M. genitalium*.

Gil et coll. ont intégré les résultats obtenus dans les études décrites précédemment (hormis Glass *et al.*, 2006) avec la recherche des gènes conservés dans les génomes de cinq endosymbiotes d'insectes (qui comportent de 544 à 621 gènes) (Gil *et al.*, 2004). A partir de cette analyse, ils proposent un génome minimal de 206 gènes comprenant 16 gènes du métabolisme ADN, 106 gènes du métabolisme ARN (dans lequel entre la traduction), 15 gènes impliqués dans la maturation et la sécrétion des protéines, 1 gène de division cellulaire, 4 gènes impliqués dans les transports (glucose et phosphate), 56 gènes du métabolisme énergétique et intermédiaire (métabolisme des lipides, des nucléotides et des cofacteurs) et 8 gènes de fonction inconnue.

Cependant, le génome minimal reste un concept puisque aucune étude jusqu'à aujourd'hui n'a pu montrer qu'une cellule avec si peu de gènes était capable de survivre. Une modélisation d'une cellule minimale a été construite avec le logiciel E-CELL, à partir de 127 gènes de *M. genitalium* (Tomita *et al.*, 1999). Néanmoins, ces essais de modélisation cellulaire restent partiels et encore peu réalistes (Takahashi *et al.*, 2003). Aujourd'hui, l'objectif est de pouvoir intégrer les données expérimentales et les données génomiques afin d'améliorer les connaissances des voies métaboliques, et de proposer un modèle plus proche de la réalité (Ishii *et al.*, 2004).

Ces études sur le génome minimal posent une autre question : quels gènes possédait la première bactérie ? Les premières bactéries étaient certainement pourvues de peu de gènes et avaient un métabolisme rudimentaire (Koonin, 2003). Peu d'études sont publiées sur l'ancêtre commun des bactéries mais cette question rejoint celle du génome minimal. Les gènes conservés dans tous les phyla bactériens proviennent peut être de l'ancêtre commun. Cependant, les génomes étant soumis à des événements de perte et d'acquisition de gènes, la recherche de gènes conservés ne saurait suffire à la caractérisation de l'ancêtre commun. Que savons-nous de l'ancêtre commun des bactéries ? Gupta et Griffiths, par l'étude de mutations dans les gènes de plusieurs génomes séquencés, ont émis l'hypothèse que la première branche à émerger a été celle des firmicutes (Gupta et Griffiths, 2002). De plus, diverses études ont montré que lors de l'évolution, les bactéries pouvaient évoluer d'une forme en bâtonnet à une forme en coque mais

que l'inverse n'est jamais observé. L'agencement des gènes *dcw* (division and cell wall) est aussi plus conservé chez les bactéries en forme de bâtonnet que chez celles en forme de coque, il est donc probable que la première bactérie ait été une bactérie en forme de bâtonnet et que la forme coque soit apparue après (Koch, 2003a). Cependant, ce sujet reste très controversé et des hypothèses existent aussi bien pour un ancêtre Gram (+) à une seule membrane que pour un ancêtre Gram (-) à deux membranes ainsi que pour les formes bâtonnet ou coque.

La génomique et la génomique comparative ont permis de mettre en évidence l'adaptation des bactéries à leur environnement. Dans un environnement idéal, les bactéries ne subiraient aucun stress et n'auraient donc pas à adapter leur physiologie en fonction des conditions. Le génome minimal obtenu par des analyses bioinformatiques ne comporte d'ailleurs aucun régulateur de transcription. Cependant, les bactéries vivent dans des environnements variables et doivent donc moduler leur physiologie selon les conditions rencontrées. L'adaptation des bactéries à leur environnement passe donc aussi par la mise en œuvre de mécanismes de résistance aux stress qu'elles sont susceptibles de rencontrer, et par une régulation fine de ceux-ci. Dans ce domaine, la génomique a permis de grandes avancées en permettant l'utilisation de techniques d'étude globale comme les puces à ADN.

2.2. *Reconstruction des réseaux de régulation*

2.2.1. *Généralités sur les systèmes de régulation transcriptionnelle*

Les bactéries, soumises à un stress, vont modifier l'expression de leurs gènes et notamment induire leurs systèmes de résistance. La mise en œuvre contrôlée de ces mécanismes passent par les systèmes de régulation. L'ensemble des gènes contrôlés par un système de régulation est appelé régulon. Il existe différents types de systèmes de régulation : les facteurs σ , les régulateurs transcriptionnels et les systèmes à deux composants.

Les facteurs σ sont des sous-unités de l'ARN polymérase qui lui permettent de reconnaître et de se fixer sur les promoteurs des gènes. Il existe différents types de facteurs σ . Le facteur σ primaire est le facteur sigma de la phase végétative. Il est essentiel à la survie des bactéries et reconnaît des séquences consensus situées à 35 (TTGACA) et 10 (TATAAT) bases en amont du site d'initiation de la transcription. L'association du σ primaire avec l'ARN polymérase est réversible. Les autres facteurs σ sont non essentiels à la survie cellulaire en condition de

croissance favorable et sont appelés facteurs σ alternatifs. Ces facteurs σ peuvent aussi interagir avec l'ARN polymérase et ainsi induire la transcription des gènes appartenant à leur régulon (Wosten, 1998).

Les régulateurs transcriptionnels sont des protéines se liant à l'ADN et permettant soit d'activer soit de réprimer la transcription du gène cible en réponse à un changement environnemental. Il existe différentes familles de régulateurs transcriptionnels qui sont définies par homologie de séquence de la région de liaison à l'ADN (composée du motif hélice-tour-hélice) (Aravind *et al.*, 2005). Certaines familles peuvent réguler un large éventail de fonctions (LysR par exemple), d'autres sont plus spécialisées (famille ArsR par exemple, spécialisée dans la réponse à des stress induits par des métaux lourds) (tableau 4). Certains régulateurs peuvent être activateur pour certains gènes et répresseur pour d'autres : les régulateurs de la famille LysR par exemple, activent en général la transcription de leurs gènes cibles et répriment leur propre expression (Schell, 1993). Les répresseurs ont leurs sites de fixation (opérateurs) dans la région promotrice du gène régulé. Une fois le répresseur fixé, l'encombrement stérique empêche l'ARN polymérase de se fixer sur le promoteur, le gène ne peut être transcrit. Un activateur fixé sur son opérateur aura un effet inverse, son interaction avec l'ARN polymérase va augmenter l'affinité de celle-ci pour le promoteur du gène régulé, qui sera transcrit de manière plus efficace.

Tableau 4 : Différentes familles de régulateurs transcriptionnels

Famille	Mode d'action		Type de fonctions régulées
	Répresseur	Activateur	
AraC		x	Catabolisme des sucres, résistance au stress et virulence
ArsR	x		Résistance aux métaux lourds
AsnC	x	x	Métabolisme des acides aminés
Crp/Fnr	x	x	Variées
DeoR	x		Catabolisme d'hydrates de carbone
GntR	x	x	Variées
LacI	x		Métabolisme carboné
LysR	x	x	Variées
LytR	x		Synthèse EPS et bactériocines, protéines extracellulaires...
MarR	x	x	Résistance aux composés toxiques, virulence, dégradation de composés aromatiques
MerR		x	Résistance aux métaux et composés toxiques
Prt	x		Biosynthèse purines et pyrimidines
RpiR	x		Systèmes PTS
TetR	x		Synthèse d'antibiotiques, résistance aux composés toxiques et au stress osmotique

Les systèmes à deux composants sont composés d'une histidine kinase et d'un régulateur de réponse. L'histidine kinase (membranaire) va sentir les stimuli extérieurs et activer par phosphorylation le régulateur de réponse. Une fois activé, le régulateur se fixe sur son site de

liaison à l'ADN et va agir de la même manière qu'un autre régulateur transcriptionnel (Bijlsma et Groisman, 2003).

Il existe d'autres systèmes de régulation indépendants de l'ARN polymérase qui font intervenir les repliements de l'ARN messenger (ARNm). Ce sont les mécanismes d'atténuation de la transcription qui agissent sur la dynamique de formation de structures « terminateur » et « antitermineur » de l'ARNm (figure 5). Les produits des gènes contrôlés par atténuation sont le plus souvent impliqués dans le transport ou le métabolisme du composé dont l'abondance influence la structure de l'ARNm. Dans les conditions où le composé est présent, la structure « terminateur » se forme et la transcription s'arrête, le gène n'est alors pas transcrit. Lorsque le composé est absent, la structure antitermineur est favorisée et les gènes sont transcrits (Merino

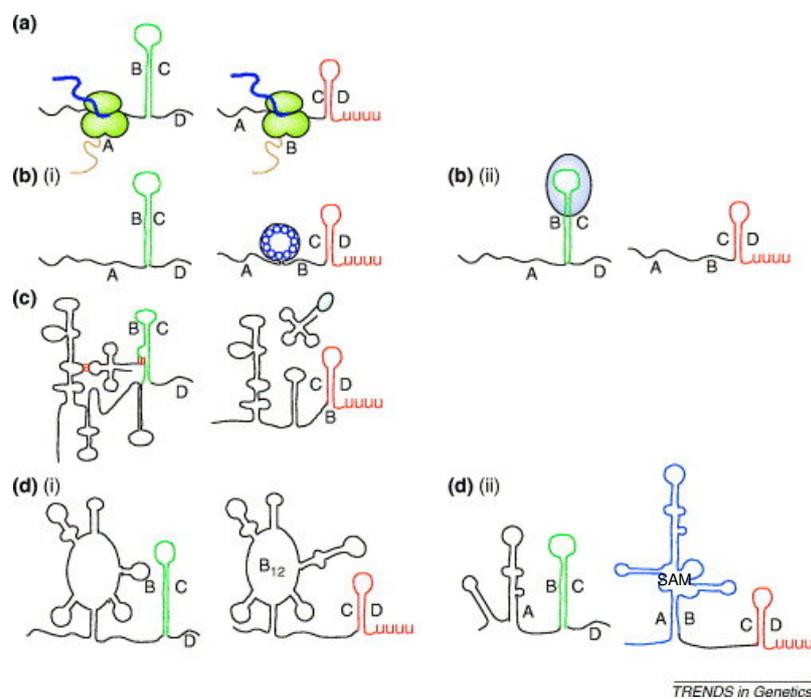


Figure 5 : Les différents mécanismes d'atténuation (Merino et Yanofsky, 2005)

Les structures anti-antitermineur (A:B), antitermineur (B:C) et terminateur (C:D) sont colorées respectivement en bleu, vert et rouge. (a) atténuation due au ribosome et au niveau d'ARNt^{trp} chargé. Si le ribosome fait une pause par manque de d'ARNt^{trp} chargé, la structure antitermineur est favorisée. (b) atténuation par fixation d'une protéine. Selon les cas, la protéine fixée peut soit favoriser la formation d'une structure terminateur (i) soit celle d'une structure antitermineur (ii). (c) Atténuation impliquant un ARNt. L'ARNt non chargé stabilise la structure antitermineur. (d) Atténuation par fixation de molécules. La molécule va induire l'atténuation de la transcription des ses gènes de biosynthèse, soit en empêchant la formation de l'antitermineur (i) soit en stabilisant la structure anti-antitermineur (ii).

et Yanofsky, 2005). Les structures terminateur ou anti-terminateur peuvent être stabilisées soit par une pause du ribosome (opéron *trp* de *E. coli*), soit par une protéine (opéron *trp* de *B. subtilis*), soit par un ARNt non chargé (T-Box des bactéries à Gram positif), soit par le composé lui-même (vitamines par exemple).

2.2.2. Identification de régulons

L'identification des réseaux de régulation et notamment des régulons de stress a fortement bénéficié des apports de la génomique. La reconstruction des réseaux de régulation nécessite le croisement de données expérimentales et de données génomiques (figure 6).

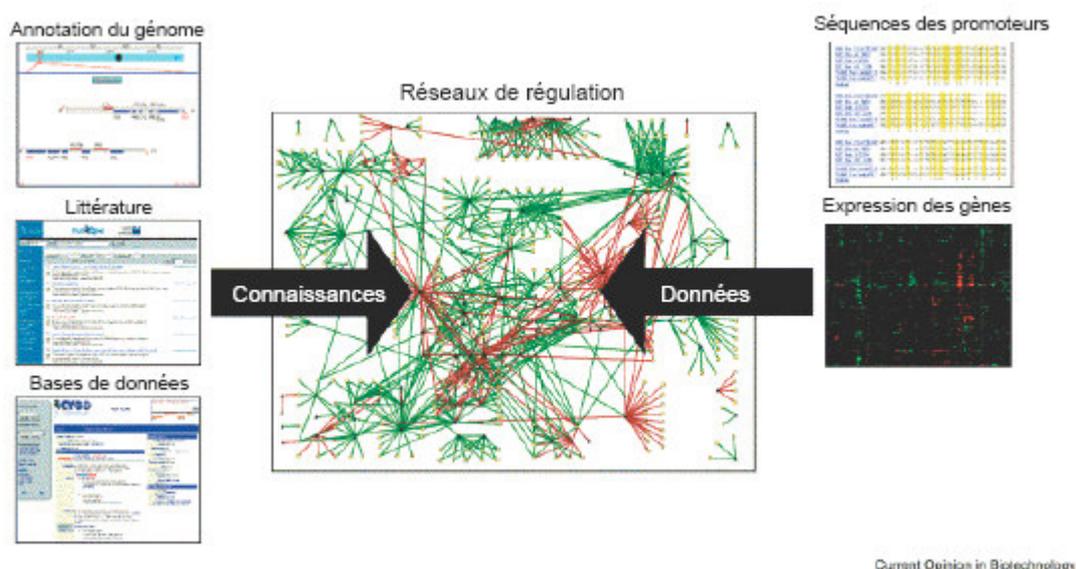


Figure 6 : Combinaison de différentes données permettant la reconstruction de réseaux de régulation (Herrgard *et al.*, 2004)

La génomique ne suffit pas à la reconstruction de réseaux de régulation. Elle nécessite le croisement de plusieurs sources de données, et notamment des données expérimentales.

L'analyse génomique permet de faire des prédictions sur les régulons comme pour l'étude du système SOS chez *Bacillus subtilis* (Au *et al.*, 2005). Le système SOS est induit par l'exposition des bactéries à des agents causant des dommages à l'ADN. Un gène est considéré comme faisant partie du système SOS (i) s'il est induit de manière dépendante de RecA par des dommages à l'ADN et (ii) s'il possède un site de liaison de LexA (répresseur du système SOS) dans son promoteur. Avant l'étude de Au *et coll.*, seuls cinq gènes remplissant ces deux critères chez *B. subtilis* étaient connus : *recA*, *lexA*, *uvrB*, *dinB* et *dinC*. Une séquence consensus du site de liaison de LexA avait déjà été publiée (Cheo *et al.*, 1991 ; Winterling *et al.*, 1998). Cette séquence a été recherchée dans l'ensemble des promoteurs putatifs du génome de *B. subtilis* et

40 sites de liaison de LexA ont été prédits. Cette prédiction a été vérifiée expérimentalement et 33 des sites prédits fixent effectivement la protéine LexA (ce qui correspond à 54 gènes sous contrôle de LexA). Une analyse transcriptomique a montré que 33 de ces 54 gènes (correspondant à 18 opérons) étaient bien induits de manière RecA dépendante lors de traitements endommageant l'ADN. Cette étude a donc permis de reconstituer le régulon SOS de *B. subtilis* à partir de prédictions informatiques (basées sur la connaissance du site de régulation) validées par des expériences d'étude globale de l'expression (puces à ADN). Sur les 33 gènes du régulon SOS identifiés, seuls 8 correspondent à des gènes du régulon SOS de *E. coli*. Cette approche consistant, à partir de données bibliographiques (séquences du site de fixation), à rechercher dans un génome des gènes candidats d'un régulon donné suivie d'une validation expérimentale a été utilisée dans différentes études. Aujourd'hui encore, l'expérimentation reste obligatoire pour valider les hypothèses fournies par l'analyse bioinformatique.

3. Les réponses aux stress

3.1. Régulation des mécanismes de résistance au stress

3.1.1. La réponse générale au stress

Les facteurs σ sont utilisés par de nombreuses bactéries pour moduler l'expression de leurs gènes en fonction de leur environnement. Parmi les facteurs σ contrôlant des régulons de réponse aux stress identifiés chez les bactéries à Gram (+), le plus étudié est probablement σ^B , qui contrôle la réponse générale au stress de *B. subtilis*. σ^B est induit en condition de stress osmotique et oxydant, en présence d'éthanol et en phase stationnaire (Boylan *et al.*, 1993). D'autres études ont montré que σ^B répond à d'autres stimuli environnementaux comme le stress hypothermique (Brigulla *et al.*, 2003) et hyperthermique (Benson et Haldenwang, 1993), le stress acide (Kovacs *et al.*, 1998) et des conditions de carence en glucose, oxygène (Volker *et al.*, 1994) ou phosphate (Volker *et al.*, 1999). L'activation du régulon σ^B permet à la bactérie d'améliorer sa résistance à différents stress (Volker *et al.*, 1999). Des analyses transcriptomiques ont montré que σ^B contrôle l'expression de plus d'une centaine de gènes (Petersohn *et al.*, 2001) en condition de stress, dont la plupart sont de fonction inconnue. Parmi les gènes régulés par σ^B se trouvent des gènes pour lesquels la relation avec la résistance aux stress est évidente comme

par exemple ceux codant pour la catalase (*katX*), la thioredoxine (*trxA*) et une protéine de liaison au fer (*dps*), impliquées dans la résistance au stress oxydant, des transporteurs de proline (*opuE*) et de glycine-bétaine (*opuD*) impliqués dans la résistance au stress osmotique ou *clpC*, codant une sous-unité des protéases Clp impliquées dans la résistance au stress hyperthermique (Hoper *et al.*, 2005). La régulation de σ^B est complexe et a été largement étudiée chez *B. subtilis*. Les protéines qui régulent l'expression de σ^B sont au nombre de sept : RsbR, RsbS, RsbT, RsbU, RsbV, RsbW et RsbU (Regulator of sigmaB). Le gène *sigB* est co-transcrit avec ces protéines régulatrices dans un opéron dont le promoteur est dépendant du sigma primaire. Lors de l'induction de σ^B , un promoteur interne dépendant de σ^B est induit, permettant l'induction des gènes *rsbV*, *rsbW*, *sigB* et *rsbX* (les quatre derniers gènes de l'opéron) (Wise et Price, 1995).

Bien que σ^B ne soit pas conservé chez tous les firmicutes (notamment chez les bactéries lactiques), des homologues de σ^B existent, par exemple chez *Listeria monocytogenes* et *Staphylococcus aureus*. Chez ces deux bactéries, σ^B est non seulement impliqué dans la réponse générale au stress mais aussi dans la virulence (Kazmierczak *et al.*, 2003 ; Bischoff *et al.*, 2004). L'organisation génétique de l'opéron *sigB* et la régulation de ce facteur σ diffèrent entre *B. subtilis* et *S. aureus*, car seuls sont présents, dans le génome de *S. aureus*, les gènes *rsbU*, *rsbV*, *rsbW* et *sigB*.

D'autres facteurs σ ont été impliqués dans la résistance aux stress mais ils contrôlent des régulons moins étendus que celui contrôlé par σ^B . Citons par exemple σ^X de *B. subtilis*, impliqué dans la résistance au stress hyperthermique (Huang *et al.*, 1997) ou σ^V de *Enterococcus faecalis* impliqué dans la survie à des stress hyperthermique, acide et à la présence d'éthanol (Benachour *et al.*, 2005).

3.1.2. Le stimulon de réponse au stress thermique : exemple de régulons coordonnés

Chez *B. subtilis*, les réponses aux stress ne dépendent pas uniquement de σ^B . Les relations entre le régulon de réponse générale au stress σ^B et la réponse à un stress spécifique peuvent être illustrée par le stimulon (ensemble des gènes induits en réponse à un stimulus) du stress hyperthermique. Les gènes de réponse au choc hyperthermique sont classés en six grandes classes chez *B. subtilis*, mais les mécanismes de régulation sont seulement connus pour les

classes I, II, III et V (Schumann, 2003). Les gènes de la classe II sont les gènes régulés par σ^B et ne sont donc pas spécifiques du stress hyperthermique.

Présentons maintenant succinctement la régulation des classes I, III et IV.

Les gènes de la classe I sont régulés par un régulateur de transcription, HrcA. Chez *B. subtilis*, ils sont organisés en deux opérons, les opérons *groES/EL* et *hrcA/grpE/dnaK/dnaJ/yqeT/yqeU/yqeV* (Homuth *et al.*, 1997). HrcA est un répresseur transcriptionnel qui se fixe sur des sites appelés CIRCE (Controlling Inverted Repeat of Chaperone Expression) de consensus : TTAGCACTC-N₉-GAGTGCTAA (Zuber et Schumann, 1994). L'activité de HrcA est modulée par la protéine GroEL (Mogk *et al.*, 1997). L'hypothèse est que, pour être active, HrcA a besoin de GroEL. Lors d'un choc hyperthermique, GroEL est titrée par les protéines dénaturées, HrcA est alors inactivé et la répression des deux opérons de chaperons est levée.

Les gènes de la classe III sont au nombre de six et sont régulés par le régulateur transcriptionnel CtsR. Ils sont organisés en trois unités transcriptionnelles, *ctsR/mcsA/mcsB/clpC*, et les unités monocistroniques *clpE* et *clpP* (Kruger et Hecker, 1998 ; Derre *et al.*, 1999a, 1999b). CtsR réprime la transcription en se fixant sur la CtsR-box dont le consensus est : ^A/_GGTCAAANAN^A/_GGTCAA. L'activité de CtsR est régulée post-traductionnellement de deux manières. CtsR est dégradée à 37°C par les protéases ATP-dépendantes ClpP et ClpX, ce qui permet de garder un niveau basal faible de CtsR (Derre *et al.*, 2000). De plus, CtsR est phosphorylée par la tyrosine kinase McsA/McsB (McsA est l'activateur de la kinase McsB), ce qui inhibe sa fixation sur la CtsR-Box (Kirstein *et al.*, 2005). L'activité de McsB est inhibée par ClpC (codée par le même opéron que CtsR, McsA et McsB). En condition de stress hyperthermique, ClpC serait titrée par les protéines dénaturées, McsB phosphorylerait alors CtsR, provoquant son inactivation ce qui lèverait la répression des gènes de la classe III.

La classe V compte deux membres, *htrA* et *htrB* codant des protéases ancrées à la membrane (Darmon *et al.*, 2002). Ces gènes sont contrôlés par le système à deux composants C_{ss}RS, qui répond non seulement au stress hyperthermique mais aussi au stress de sécrétion (induit en laboratoire par surproduction d' α -amylase). C_{ss}S détecterait l'accumulation de protéines dénaturées dans le milieu extérieur et activerait, via la phosphorylation de C_{ss}R et sa fixation sur les opérateurs, l'expression des protéases de surface (Hyrylainen *et al.*, 2001).

La régulation du stress hyperthermique chez *B. subtilis* fait donc intervenir différents types de régulateurs. De plus, certains gènes sont sous double contrôle, comme les gènes de classe III, *clpC* et *clpP* sous contrôle de σ^B et de CtsR.

Chez les autres firmicutes, l'organisation observée chez *B. subtilis* n'est pas toujours conservée. Par une étude de génomique comparative, Chastanet et coll. ont montré que les sites de liaison de CtsR se retrouvent en amont des opérons *groESL* des streptocoques (Chastanet *et al.*, 2003a) et cette régulation a été démontrée chez *S. pneumoniae* (Chastanet *et al.*, 2001). Chez *Streptococcus salivarius*, *clpP* est sous contrôle de CtsR et d'HrcA (Chastanet et Msadek, 2003b). Chez *Staphylococcus aureus*, CtsR contrôle non seulement les gènes *clpC*, *clpB* et *clpP* mais aussi les opérons *dnaK* et *groESL*, qui se retrouvent sous le double contrôle de CtsR et HrcA (Chastanet *et al.*, 2003a). Chez la bactérie lactique *Oenococcus oeni*, seul CtsR est présent, HrcA est absent de la séquence du génome. Chez cette bactérie, les gènes *clp* et les opérons *groESL* et *dnaK* sont sous le contrôle unique de CtsR (Grandvalet *et al.*, 2005, figure 7).

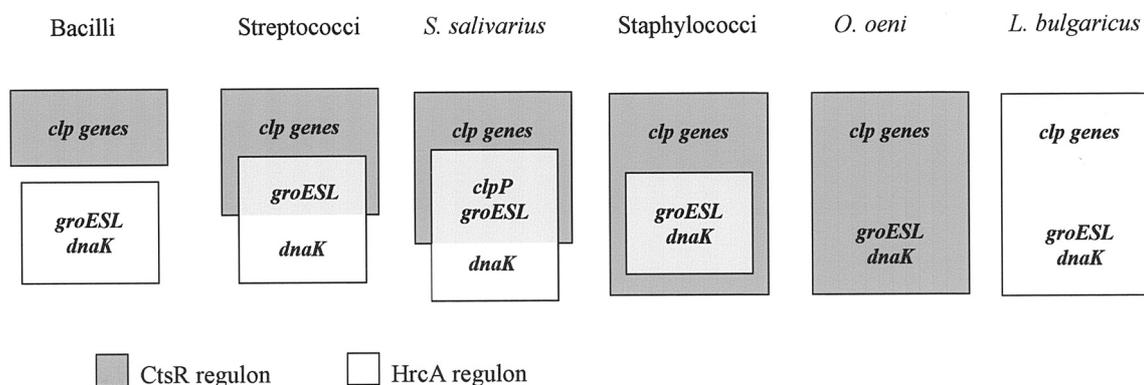


Figure 7 : Comparaison des régulons d'HrcA et CtsR chez différents firmicutes
(Grandvalet *et al.*, 2005)

3.1.3. Les BL et la régulation des réponses aux stress

Chez les bactéries lactiques, peu de régulons de stress sont caractérisés. La plupart des études menées jusqu'à aujourd'hui sont des études globales permettant d'identifier des gènes induits ou des protéines surproduites en condition de stress. Cependant, ces études ne se sont généralement pas focalisées sur la régulation des mécanismes de résistance mis en jeu. Les

génomomes de bactéries lactiques dont la séquence a été publiée montrent que ces bactéries possèdent en général peu ou pas de facteur σ alternatif et notamment pas d'homologue de σ^B .

Parmi les systèmes de régulation, les systèmes à deux composants semblent de bons candidats pour la régulation coordonnée de la réponse aux changements environnementaux. Certaines études ont été réalisées par l'inactivation spécifique de ces systèmes chez *L. lactis* (O'Connell-Motherway *et al.*, 2000) et chez *L. sakei* (Morel-Deville *et al.*, 1998). Chez *L. lactis*, l'inactivation d'un des systèmes à deux composants induit la perte de l'activité de la voie arginine déiminase et une augmentation de la sensibilité au stress acide. Chez *L. sakei*, certains mutants sont plus sensibles à certains stress comme le stress hyperthermique, acide ou oxydant. Cependant, dans les deux cas, les régulons des systèmes à deux composants étudiés ne sont pas déterminés. Récemment, une étude portant sur un système à deux composants de *L. acidophilus* a été réalisée en transcriptomique (Azcarate-Peril *et al.*, 2005). Le mutant est plus sensible au stress acide et la comparaison entre la souche sauvage et le mutant montre que la plupart des gènes différentiellement exprimés appartiennent aux systèmes de protéolyse et de transport d'acides aminés ou de peptides. Cependant, encore une fois, le mécanisme d'activation de ce système à deux composants est inconnu et rien ne permet de conclure à une action directe sur les promoteurs de ces gènes ou à une action indirecte.

Les bactéries lactiques rencontrent, à priori, au cours des procédés industriels, dans leur environnement ou lors de la traversée du transit digestif *in vivo*, de nombreux stress, comme le stress thermique (hyper ou hypothermique), le stress osmotique, le stress oxydant ou le stress biliaire. De part leur métabolisme fermentaire, les bactéries lactiques sont soumises au stress acide, stress auquel nous nous sommes plus particulièrement intéressés.

3.2. Le stress acide

3.2.1. Effets de l'acide lactique sur les bactéries

Les bactéries lactiques produisent au cours de la croissance de l'acide lactique qui inhibe pour partie le développement de la flore indésirable dans les produits fermentés (Shelef, 1994). L'acide lactique est un acide faible dont le pKa est de 3,86. Il existe sous deux formes : la forme non dissociée $\text{CH}_3\text{CHOHCOOH}$ (acide lactique) et la forme dissociée $\text{CH}_3\text{CHOHCOO}^-$ (lactate). L'équilibre entre les deux formes dépend du pH de la solution : plus le pH est bas, plus la forme non dissociée est représentée. Sous sa forme non dissociée, l'acide lactique peut diffuser à travers la membrane. Le pH intracellulaire étant plus élevé que le pH externe (pour les conditions de croissance à pH neutre ou acide), l'acide lactique entrant se dissocie en lactate et libère un proton (Russell et DiezGonzalez, 1998). Le lactate, molécule chargée, ne diffuse pas à travers la membrane et s'accumule dans la cellule. L'entrée d'acide lactique se faisant par diffusion passive, celle-ci s'arrête lorsque la concentration interne en acide lactique est égale à celle de l'extérieur. Comme l'équilibre entre les formes dissociée et non dissociée dépend du pH, l'entrée passive d'acide lactique dans la cellule est dépendante du ΔpH et se fait tant que le pH interne est supérieur au pH externe.

L'acide lactique induit donc l'acidification du cytoplasme et l'accumulation de lactate dans la cellule. Outre les effets délétères de l'acidification du cytosol, l'accumulation de lactate dans la cellule peut conduire à l'inhibition de la fermentation lactique, à une perte d'activité de l'eau et à la chélation d'éléments nécessaires (Presser *et al.*, 1997).

3.2.2. Mécanismes de résistance au stress acide des BL

Chez les bactéries neutrophiles et notamment chez les bactéries lactiques, l'incubation à un pH sublétal (pH d'adaptation) améliore la résistance à une épreuve acide létale : ce phénomène est appelé « Acid Tolerance Response » (ATR). Cette ATR a été décrite chez différents lactobacilles comme *L. sanfranciscensis* (De angelis *et al.*, 2001), *L. collinoides* (Laplace *et al.*, 1999), *L. acidophilus* (Lorca *et al.*, 1998), *L. plantarum* (McDonald *et al.*, 1990) et *L. bulgaricus* (Lim *et al.*, 2000) ainsi que pour des lactocoques, streptocoques et enterocoques (pour revue : van de Guchte *et al.*, 2002). Les mécanismes de résistance au stress acide décrits peuvent être classés en différentes catégories (Cotter et Hill, 2003) : les mécanismes de maintien de la force protomotrice, ceux permettant la production de composés alcalins, les modifications

de l'enveloppe cellulaire, les mécanismes de prise en charge des protéines dénaturées et les mécanismes de réparation de l'ADN.

Dans cette partie, nous verrons les mécanismes de résistance au stress acide mis en évidence chez les lactobacilles et qui pourraient se retrouver dans le génome de *L. bulgaricus*. Lorsque les données ne sont pas disponibles pour les lactobacilles, nous nous attacherons aux résultats obtenus pour les autres BL.

3.2.2.1. *Maintien de la force protomotrice*

Le gradient électrochimique de protons à travers la membrane ($\Delta\mu\text{H}^+$ ou force protomotrice) est la résultante de deux composantes : le potentiel électrique ($\Delta\Psi$) (milieu intracellulaire plus négatif que le milieu externe) et le gradient de protons (ΔpH) (compartiment intracellulaire plus alcalin que le milieu externe). La relation liant l'ensemble est : $\Delta\mu\text{H}^+ = \Delta\Psi - Z\Delta\text{pH}$, Z étant une constante dépendant de la température ($Z = 59$ mV à 25°C par exemple) (Kakinuma, 1998). Chez les bactéries possédant une chaîne respiratoire, celle-ci crée un gradient de protons à travers la membrane en les expulsant de la cellule. L'énergie créée par ce gradient de protons est utilisée par la F_1F_0 -ATP synthase (ou H^+ -ATPase) pour la synthèse de l'ATP à partir d'ADP et de phosphate inorganique.

De manière plus générale, un certain nombre de transports demandant de l'énergie est lié aux mouvements de protons à travers la membrane, et par là-même à la force protomotrice. Si le transport se fait dans le même sens que le mouvement de protons à travers la membrane, on parle de symport. Si le transport se fait dans le sens inverse, on parle alors d'antiport (figure 8) (Kakinuma, 1998).

Pendant longtemps, les BL ont été considérés comme dépourvus de chaînes respiratoires. Il a été établi qu'en présence d'une source d'hème (et parfois de quinone), certaines BL sont capables de respirer, ce qui se traduit par une acidification restreinte et un gain de viabilité en phase stationnaire (*L. lactis*, Duwat *et al.*, 2001 ; Rezaiki *et al.*, 2004 ; *E. faecalis*, Winstedt *et al.*, 2000 ; streptocoques, Yamamoto *et al.*, 2005). Ces travaux ont conduit à une nouvelle méthode de préparation des levains et ferments mais la biotransformation des aliments continue de se faire en condition de fermentation, la production d'acide lactique étant la plus souvent recherchée. Lors du processus de fermentation, la chaîne respiratoire des BL est inactive, elle ne peut donc pas participer à la génération de la force protomotrice.

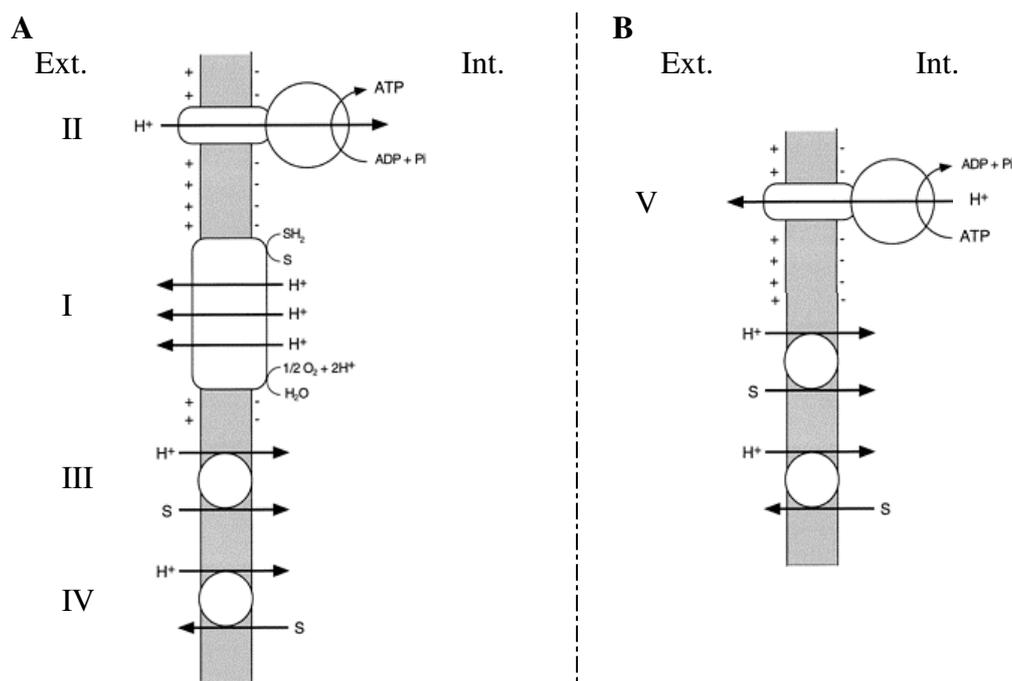


Figure 8 : Couplage de la force protomotrice avec des cotransports de protons

A- L'expulsion électrogénique de H^+ par la chaîne respiratoire (I) permet d'établir un $\Delta\mu H^+$ qui sera ensuite utilisé pour produire de l'ATP (II) et permettre le transport actif de différents substrats (S) par des systèmes de symport (III) ou d'antiport (IV).

B- Lorsque la chaîne respiratoire est inactive (par exemple en condition de fermentation), l'expulsion de protons est réalisée par la H^+ -ATPase (V) qui va établir le $\Delta\mu H^+$ en utilisant de l'ATP. Dans certains cas, ce processus peut être inversé (voir dans le texte).

Adapté de Kakinuma, 1998.

L' H^+ -ATPase

Les premières études sur le maintien du pH intracellulaire par l' H^+ -ATPase ont été menées chez *E. faecalis*. Les travaux de Harold et coll. ont montré que chez cette bactérie, l' H^+ -ATPase est responsable de l'expulsion des protons couplée à la dépense d'ATP (Harold *et al.*, 1970). D'autres travaux, notamment ceux de Kobayashi, ont permis de confirmer que l' H^+ -ATPase a un rôle majeur dans l'homéostasie du pH cytoplasmique (Kobayashi *et al.*, 1982 ; Kobayashi, 1985 ; Kobayashi *et al.*, 1986). Un mutant de l' H^+ -ATPase ne maintient plus de ΔpH lorsque le pH diminue et sa croissance est très ralentie à pH 6,0 (Kobayashi, 1985). La régulation du pH intracellulaire est fonction de la quantité et de l'activité de la H^+ -ATPase (Kobayashi *et al.*, 1986). A pH acide, la quantité des sous-unités α et β de H^+ -ATPase augmente et elle diminue

lorsque les cellules sont incubées à pH légèrement alcalin. L'alcalinisation du milieu intracellulaire diminue vers pH 8, pH auquel la H⁺-ATPase n'est plus active. L'expulsion de protons est un phénomène électrogénique (elle augmente le potentiel de membrane par l'expulsion de charges positives).

Chez *L. lactis*, des études ont montré que la H⁺-ATPase avait un rôle dans l'ATR car l'activité de la H⁺-ATPase augmente avec la tolérance à l'acidité (O'sullivan et Condon, 1999). Cette enzyme est essentielle à la croissance de *L. lactis* (Koebmann *et al.*, 2000). Plus récemment, des études sur des lactobacilles probiotiques (*L. rhamnosus* GG et E800, *L. paracasei* NFBC 338, *L. salivarius* UCC 500 et *L. gasseri* ATCC 33323) ont montré que leur survie à un environnement acide était augmentée en présence de sucres métabolisables et notamment de glucose. L'explication fournie est que le glucose permet de maintenir une activité glycolytique générant de l'ATP (mais aussi de l'acide lactique) qui sera ensuite utilisé par la H⁺-ATPase pour augmenter le pH intracellulaire (Corcoran *et al.*, 2005).

Autres ATPases

D'autres types d'ATPases pourraient être impliqués dans l'homéostasie du pH. L'expulsion électrogénique de protons par l'H⁺-ATPase augmente le potentiel de membrane. Pour contrer cet effet, des transports d'ions sont impliqués et notamment le transport d'ions K⁺ (Dashper et Reynolds, 1992 ; Kashket et Barker, 1977 ; Kashket *et al.*, 1980). Ce transport d'ions K⁺ nécessite chez *E. hirae* la consommation d'ATP (Bakker et Harold, 1980).

Plus récemment une ATPase de type P présentant les caractéristiques des H⁺-ATPases de levure a été étudiée chez *S. mutans*. Cette ATPase, caractérisée biochimiquement, pourrait transporter soit des protons, soit médier un antiport ion/H⁺ et serait impliquée dans la régulation du pH intracellulaire (Magalhaes *et al.*, 2003, 2005).

Décarboxylation d'acides aminés

L'hypothèse d'un rôle des décarboxylases d'acides aminés dans la résistance au stress acide remonte aux années 1920 (Hanke et Koessler, 1924 ; Gale, 1940), suite à l'observation que l'activité décarboxylase augmentait avec la baisse du pH chez des bactéries isolées de fécès, notamment *E. coli*. La décarboxylation d'un acide aminé consomme un proton intracellulaire et produit l'amine correspondant à l'acide aminé et du CO₂. De nombreuses décarboxylases ont été étudiées chez les BL. En 1952, Lagerborg et Clapper testent la capacité de 33 souches de lactobacilles (31 souches isolées de salive humaine, *L. casei* ATCC 7469 et *L. arabinosus* ATCC

8014) à produire du CO₂ à partir de différents acides aminés (arginine, tyrosine, lysine, histidine, ornithine, glutamate). Ils montrent que dans cet échantillon de souches, les réactions les plus fréquentes sont : la décarboxylation de l'arginine (12/33), celle du glutamate et de la tyrosine n'ont lieu que dans 3 souches sur 33, et la décarboxylation de l'histidine et de l'ornithine dans une seule souche. Les décarboxylases ne sont donc pas systématiquement présentes chez les lactobacilles et leur présence varie selon les souches (les souches de *L. casei* ATCC 7469 et *L. arabinosus* ATCC 8014 ne possèdent aucune des décarboxylases).

La première décarboxylase étudiée pour son rôle dans la résistance au stress acide chez les lactobacilles a été l'histidine décarboxylase (Recsei et Snell, 1972). Cette enzyme produit de l'histamine et du CO₂ à partir de l'histidine. Lorsque *Lactobacillus* 30a est placé dans un milieu à pH 4,3 en présence d'histidine, il produit du CO₂ et de l'histamine, le pH du milieu augmente jusqu'à pH 4,8 et la souche se développe. Dans le même milieu, un mutant dépourvu d'activité histidine décarboxylase est incapable d'augmenter le pH et croît peu. Chez *L. buchneri* l'histamine produite au cours de la décarboxylation de l'histidine est exportée par un antiport électrogénique Histidine⁺/Histamine⁺ permettant la création d'un $\Delta\Psi$, qui, en association avec le ΔpH créé lors de la réaction de décarboxylation, génère une force protomotrice (Molenaar *et al.*, 1993, figure 9). La régulation de l'activité de l'histidine décarboxylase a été étudiée chez *Lactobacillus* 30a. Le pH induit des changements dans la conformation de l'enzyme qui stabilisent l'hélice du site de liaison du substrat et permettent à l'enzyme d'être active à pH acide (Schelp *et al.*, 2001).

D'autres décarboxylases ont été étudiées. Chez *L. brevis* (Ueno *et al.*, 1997), *Lactobacillus* E1 (Higuchi *et al.*, 1997) et *L. lactis* (Sanders *et al.*, 1998), une glutamate décarboxylase a été mise en évidence, couplée à un antiport Glutamate⁻/GABA. Chez *L. lactis*, l'induction du système décarboxylase-antiport dépend du pH, de la présence de glutamate et d'ions Cl⁻ (Sanders *et al.*, 1998). Chez *L. lactis*, l'inactivation de ce système induit une sensibilité accrue au stress acide (Sanders *et al.*, 1998). *L. brevis* présente également une tyrosine décarboxylase dont le rôle potentiel dans la résistance à l'acidité n'a pas été étudié (Lucas *et al.*, 2003).

Il a été également proposé que les décarboxylases aient un autre rôle qui contribue à la tolérance à l'acidité, la génération d'une force protomotrice. Chez *Lactobacillus* M3, l'ATP cellulaire augmente fortement en présence d'aspartate (Abe *et al.*, 1996) mais la présence d'inhibiteur de la H⁺-ATPase ou d'ionophores inhibe la synthèse d'ATP. Dans cette souche, la présence d'un système aspartate décarboxylase -antiport Aspartate⁻/Alanine a été mise en

évidence. La réaction de décarboxylation couplée au transport électrogénique permettrait de générer une force motrice qui serait utilisée par la H^+ -ATPase pour produire de l'ATP.

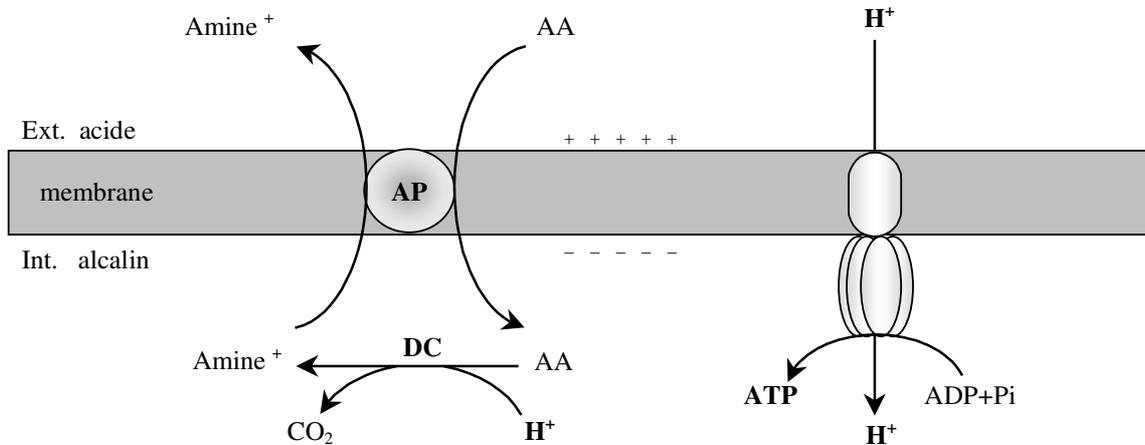


Figure 9 : Modèle de la création de la force motrice par la décarboxylation d'acides aminés et de son utilisation pour la synthèse d'ATP

L'acide aminé (AA) est introduit dans la cellule par un système d'antiport (AP) et dégradé pour produire l'amine correspondante et du CO_2 par la décarboxylase (DC). Cette réaction consomme un proton et crée ainsi un ΔpH . L'amine ($Amine^+$) qui possède une charge positive supplémentaire par rapport à l'AA est expulsée par le système antiport contre l'import de l'acide aminé. Ce transport est électrogénique et crée un $\Delta\Psi$ qui se répercute sur la force motrice. Le $\Delta\mu H^+$ ainsi créé peut être utilisé par l' H^+ -ATPase pour produire de l'ATP.

Fermentation malolactique

Un autre transport électrogénique, couplé à la fermentation malolactique a aussi été mis en évidence chez les BL. La fermentation malolactique est la conversion de l'acide malique (acide dicarboxylique) en acide lactique (acide monocarboxylique). Cette conversion consomme un proton et libère du CO_2 . Elle a surtout été étudiée chez *Oenococcus oeni* et *L. lactis* (Renault *et al.*, 1988). Les premières observations ont montré que la croissance d'*O. oeni* était stimulée par la présence d'acide malique dans le milieu, notamment à bas pH (Pilone et Kunkee, 1976). Les transporteurs responsables de la création du $\Delta\mu H^+$ sont de deux types : un antiport Malate²⁻/Lactate⁻ présent chez *L. lactis* (Poolman *et al.*, 1991) et *L. sakei* (Champomier-Vergès *et al.*, 2001) et un uniport Malate⁻ chez *O. oeni* (Salema *et al.*, 1994), *L. plantarum* (Olsen *et al.*, 1991), *L. sakei* (Champomier-Vergès *et al.*, 2001) et *L. lactis* (Poolman *et al.*, 1991). Les deux types de transport sont électrogéniques et contribuent au $\Delta\Psi$. La fermentation malolactique consommant un proton, elle contribue au ΔpH . Le $\Delta\mu H^+$ généré peut être utilisé par la cellule pour produire de l'ATP. En 1989, Cox et Henick-Kling montrent qu'à pH 3,5 et 5,5, la présence de malate augmente la quantité d'ATP intracellulaire. Cette augmentation d'ATP est inhibée par l'addition

d'inhibiteur de l' H^+ -ATPase (DCCD) ou d'ionophore (monensine), ce qui indique qu'elle dépend du $\Delta\mu H^+$.

Fermentation citrolactique

Certaines BL sont capables d'utiliser le citrate comme source de carbone via la fermentation citrolactique. Le métabolisme du citrate nécessite plusieurs enzymes dont les gènes sont détectés dans plusieurs genres comme *Weissella*, *Leuconostoc*, *Lactobacillus*, *Oenococcus* et *Lactococcus* (Martin *et al.*, 2005). La première étape est le clivage du citrate en acétate et oxaloacétate par le complexe citrate lyase, codé par les gènes *citCDEFG* (CitD, E, F sont les sous-unités du complexe et citG et citC sont nécessaires à son activité). L'oxaloacétate est ensuite décarboxylé en pyruvate par l'oxaloacétate décarboxylase codée par le gène *citM*. La décarboxylation consomme un proton (permettant l'alcalinisation du milieu intracellulaire et générant un ΔpH) et produit du CO_2 . Le pyruvate sera ensuite transformé en lactate par la lactate déshydrogénase. Le lactate ainsi produit est ensuite excrété par un système d'antiport Lactate⁻/Citrate²⁻ électrogénique permettant la création d'un $\Delta\Psi$. La combinaison du ΔpH et du $\Delta\Psi$ crée une force protomotrice qui sera utilisée par la cellule pour la synthèse d'ATP (Marty-Teysset *et al.*, 1996 ; Drider *et al.*, 2004). La relation entre la fermentation citrolactique et la résistance au stress acide a été étudiée chez *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis*. Chez cette bactérie, les fonctions de transport et les enzymes de dégradation du citrate sont codées par deux opérons différents, situés respectivement sur un plasmide et sur le chromosome (Martin *et al.*, 2004). Les deux opérons sont induits par le stress acide, indépendamment de la présence de citrate (Garcia-Quitans *et al.*, 1998 ; Martin *et al.*, 2004).

3.2.2.2. *Production de composés alcalins*

Un autre moyen de contrôler le pH lors de l'acidification du milieu est de produire des composés alcalins comme l'ammoniac (NH_3). Le NH_3 contribue à alcaliniser le milieu en réagissant avec un proton pour former du NH_4^+ . Chez les BL, deux voies enzymatiques produisant du NH_3 sont connues : la voie de l'arginine déiminase et l'uréase.

La voie de l'arginine déiminase

La voie de l'arginine déiminase (ADI) a été identifiée chez de nombreuses BL, elle n'est cependant pas présente chez toutes (Manca de Nadra *et al.*, 1988 ; Cunin *et al.*, 1986). A titre d'exemples, l'ADI est présente chez les lactobacilles hétérofermentaires du vin mais pas chez les lactobacilles homofermentaires issus de ce même milieu (Liu *et al.*, 1995 ; Liu et Pilone, 1998).

Cependant, certains lactobacilles homofermentaires, *L. leichmannii* par exemple (Hutson et Downing, 1968), ont la capacité de dégrader l'arginine par cette voie.

L'ADI conduit à la dégradation de l'arginine en ammoniac, ornithine et CO₂ et est composée de trois enzymes qui catalysent les réactions enzymatiques suivantes:

Arginine déiminase (ADI) Arginine + H₂O → Citrulline + NH₃

Ornithine transcarbamylyase (OTC) Citrulline + P_i → Ornithine + Carbamyl-P_i

Carbamate kinase (CK) Carbamyl-P_i + ADP + H₂O → ATP + NH₃ + CO₂

L'ornithine produite est transportée hors de la cellule par un antiport Arginine/Ornithine électroneutre (Champomier-Verges *et al.*, 1999). Cette voie de dégradation de l'arginine permet donc non seulement d'augmenter le pH intracellulaire par production de NH₃, mais aussi de produire de l'énergie sous forme d'ATP. Les effets de l'ajout d'arginine dans le milieu de culture ont surtout été étudiés sur la survie en phase stationnaire. La présence d'arginine augmente la survie et le pH de fin de culture de *L. sakei* notamment en condition d'anaérobiose (Champomier-Verges *et al.*, 1999), de *L. sanfranciscensis* (De Angelis *et al.*, 2002), de *L. buchneri* (Mira De Orduna *et al.*, 2001) et de *L. reuteri* (Rollan *et al.*, 2003).

Le rôle du catabolisme de l'arginine sur la résistance au stress acide et l'ATR a été étudié chez *L. lactis* (Budin-Verneuil *et al.*, 2004) et chez *L. reuteri* (Rollan *et al.*, 2003). Chez *L. reuteri*, la présence d'arginine lors de l'adaptation à l'acidité est indispensable à la mise en place de l'ATR. Chez ce lactobacille, cette voie serait donc une composante essentielle de l'ATR. Ce n'est cependant pas un rôle général de l'ADI. Ainsi chez *L. sakei*, l'effet bénéfique sur la survie en phase stationnaire n'est pas lié à l'effet pH mais plutôt à l'énergie produite sous forme d'ATP (Champomier-Verges *et al.*, 1999).

La régulation de cette voie a été étudiée chez plusieurs lactobacilles et les résultats divergent selon les espèces. Chez tous les lactobacilles étudiés, l'ADI est induite en présence d'arginine. Cependant, chez certains lactobacilles, elle est aussi soumise à la répression catabolique (*L. sakei* (Zuniga *et al.*, 1998) ou *L. reuteri* (Rollan *et al.*, 2003)) alors que pour d'autres, elle échappe à ce contrôle (*L. sanfranciscensis* (De Angelis *et al.*, 2002), *L. plantarum* (Spano *et al.*, 2004) ou *L. buchneri* (Manca de Nadra *et al.*, 1988 ; Mira de Orduna *et al.*, 2001)). De plus, chez *L. sakei*, l'ADI est induite en anaérobiose, peut être du fait du contrôle d'un régulateur de type Crp/Fnr indispensable à l'expression des gènes de cette voie (Zuniga *et al.*, 2002).

L'uréase

L'uréase permet la production de NH_3 à partir d'urée. Elle a été mise en évidence chez *S. thermophilus* (Mora *et al.*, 2004, 2005) et *L. fermentum* (Suzuki *et al.*, 1979) mais a surtout été étudiée chez *S. salivarius*. L'uréase est une enzyme multimérique nécessitant du nickel pour son activité catalytique. Elle dégrade l'urée ($(\text{NH}_2)_2\text{CO}$) en CO_2 et 2NH_3 . Chez *S. salivarius*, l'uréase a deux rôles principaux : elle permet à la bactérie de résister à des pH acides et d'utiliser l'urée comme source d'azote (Chen *et al.*, 2000).

3.2.2.3. Modifications de l'enveloppe cellulaire

L'enveloppe cellulaire constitue l'interface entre l'environnement et le milieu intracellulaire. Elle est le siège des transports et est la plus exposée aux stress extracellulaires. Il n'est donc pas étonnant qu'elle joue un rôle majeur dans l'adaptation aux changements environnementaux.

Chez plusieurs espèces de BL, des modifications de la composition en acides gras de la membrane ont été observés en réponse aux changements du pH externe. Chez *S. sanguis*, *S. gordonii*, *S. salivarius* et *L. casei*, le rapport entre $\text{C}_{18:1} + \text{C}_{20:1}$ et $\text{C}_{14:0} + \text{C}_{16:0}$ augmente drastiquement lors de l'acidification du milieu (Fozo *et al.*, 2004a). Chez *S. mutans*, l'enrichissement de la membrane en acides gras insaturés à longues chaînes $\text{C}_{18:1}$ et $\text{C}_{20:1}$ diminuerait la perméabilité de la membrane aux protons et augmenterait la résistance aux pH acides (Quivey *et al.*, 2000). En présence d'un inhibiteur de la synthèse des acides gras (céruléine), *S. mutans* ne développe plus d'ATR malgré une incubation au pH d'adaptation (pH 5) (Fozo et Quivey, 2004b). Ce résultat établit que le changement de composition membranaire aurait un rôle important pour la résistance à l'acidité.

Fozo et coll. (2004c) ont identifié l'enzyme de *S. mutans* responsable de la désaturation des acides gras, FabM. (analogue de FabA de *E. coli*). La déficience en FabM augmente la perméabilité membranaire aux protons et rend la souche plus sensible au stress acide (Fozo *et al.*, 2004c).

Chez *E. coli*, l'enzyme responsable de la formation des acides gras cyclopropaniques (Cfa synthase) est impliquée dans la résistance au stress acide et notamment dans la mise en place de l'adaptation (Chang et Cronan, 1999). Cependant, le phénotype du mutant *cfa* est surtout visible dans un mutant *rpoS* (facteur σ de phase stationnaire). Chez *Lactococcus lactis*, la *cfa* synthase est fortement induite transcriptionnellement durant l'ATR (Budin-Verneuil *et al.*, 2005a). Cependant, l'implication de la Cfa synthase dans la résistance à l'acidité ou l'ATR n'a pas été démontrée.

Plusieurs autres fonctions impliquées dans le métabolisme de l'enveloppe affectent la résistance à l'acidité de *S. mutans*.

Le gène *dagK* est impliqué dans la biogenèse des acides phosphatidiques membranaires. L'inactivation de ce gène rend la souche incapable de croître en dessous de pH 5,5 (Yamashita *et al.*, 1993). *DagK* pourrait aussi être impliqué dans la transduction de signaux environnementaux.

Ffh fait partie de la machinerie de sécrétion des protéines néosynthétisées appelée particule de reconnaissance du signal (SRP). SRP est constituée d'une protéine chaperon (*Ffh*) et d'un petit ARN cytoplasmique (Koch *et al.*, 2003b). Chez *S. mutans*, une mutation du gène *ffh* induit une sensibilité accrue au stress acide (Gutierrez *et al.*, 1999). Chez ce mutant, en condition d'adaptation, l'activité de la H⁺-ATPase et l'ATR ne sont plus induites. Une étude plus récente révèle qu'en culture continue à pH 5 l'activité de la H⁺-ATPase est bien induite chez *ffh* mais que le niveau d'activité de l'enzyme est très inférieur à celui de la souche sauvage (Crowley *et al.*, 2004). Cette faible activité pourrait s'expliquer par un défaut d'assemblage de l'H⁺-ATPase dans la membrane. Dans la mesure où une mutation *ffh* a probablement un effet pléiotrope, d'autres protéines impliquées dans l'ATR pourraient également être affectées.

La paroi est, elle aussi, impliquée dans la résistance à un environnement acide. Chez *S. mutans*, l'inactivation de *dltC* codant la protéine D-alanyl-carrier (permettant l'incorporation de D-alanine dans les acides lipotéchoïques associés à la membrane) conduit à des défauts de croissance et à une forte diminution de la tolérance à l'acidité, et notamment à une forte diminution de l'ATR (Boyd *et al.*, 2000a). La modification de la paroi due à l'inactivation de *dltC* semble augmenter la perméabilité de la cellule aux protons, et ainsi la rendre plus sensible au stress acide, notamment en empêchant la mise en place de l'ATR.

3.2.2.4. *Prise en charge des protéines dénaturées*

Les protéines chaperons ont un rôle très important dans la cellule. Elles sont nécessaires au bon repliement des protéines néosynthétisées ou dénaturées par un stress et elles limitent les phénomènes d'agrégation protéique (Craig *et al.*, 1993). En 1997, des études chez *L. lactis* montrent pour la première fois que les protéines chaperons GroES et GroEL ne sont pas seulement induites par un stress thermique mais le sont aussi à pH d'adaptation (pH 5,5) (Hartke *et al.*, 1997). Les études menées chez *L. bulgaricus* (Lim *et al.*, 2000) montrent que lors de la mise en place de l'ATR, GroES, GroEL et DnaK sont induites. Cependant il semble que

l'induction des protéines chaperons ne soit pas une règle générale. Chez *S. mutans* (Wilkins *et al.*, 2001) et *S. oralis* (Wilkins *et al.*, 2002) seul GroEL est induit à pH 5,2. D'autres études chez *S. mutans* montrent une induction transitoire de GroEL et DnaK lors du passage de pH 7 à pH 5 puis un retour au niveau initial en condition de culture prolongée à pH 5 (Lemos *et al.*, 2001). Cependant, des résultats contradictoires ont été obtenus par Len et coll. qui montrent que DnaK et GroEL sont induits en condition de culture à pH acide contrôlé (pH 5) par rapport à une culture à pH 7 (Len *et al.*, 2004). Chez *L. sanfranciscensis*, aucune différence d'expression de GroELs et DnaK n'est détectable entre des cultures adaptées (1h à pH 5) et des cultures non adaptées (pH 6,4) (De Angelis *et al.*, 2001). L'induction des protéines chaperons, décrites dans de nombreuses bactéries, n'a cependant pas été démontrée comme nécessaire à la résistance au stress acide ou à l'ATR.

L'étude de la réponse au stress thermique a montré que lorsque les protéines ne peuvent pas être renaturées, elles sont dégradées par les protéases ATP-dépendantes de la famille Clp, ce qui permet de recycler leurs acides aminés. La protéolyse par Clp requiert une peptidase de type serine (la sous-unité ClpP) et une sous-unité ATPase régulatrice (Gottesman *et al.*, 1997). Seules, les sous-unités ATPase remplissent un rôle de chaperon protéique, qu'elles n'exercent plus lorsqu'elles interagissent avec ClpP (pour revue : Houry, 2001). Chez *L. lactis*, *clpP* est induit lors de l'ATR (Frees et Ingmer, 1999). Chez *S. mutans*, *clpP* et *clpC* sont induits lors de l'ATR et un mutant *clpP* présente un phénotype pléiotrope : croissance ralentie, formation de longues chaînettes et sensibilité à différents stress dont le stress acide (Lemos et Burne, 2002).

HtrA est une protéase extracellulaire très conservée chez les procaryotes. Elle est impliquée dans le contrôle de la qualité des protéines en dégradant les protéines anormales exportées et dans la maturation des protéines sécrétées (Poquet *et al.*, 2000). Chez *S. mutans*, un mutant HtrA croît moins bien à pH acide que la souche sauvage mais les bases moléculaires de ce phénotype ne sont pas établies (Diaz-Torres et Russell, 2001).

3.2.2.5. Réparations de l'ADN

L'acidification du cytosol est susceptible de conduire à l'apparition de sites abasiques dans l'ADN (Lindahl et Andersson, 1972). Chez *S. mutans*, un mutant *recA* est plus sensible à un stress acide létal. Cependant, lorsque l'ATR est induite par incubation à pH 5, la survie du mutant et de la souche sauvage à un stress létal est identique (Quivey *et al.*, 1995). RecA interviendrait donc dans le processus de survie au stress acide, mais pas dans le processus d'adaptation.

Cependant, l'observation qu'une incubation à pH 5 permet d'augmenter la tolérance aux radiations UV et au peroxyde d'hydrogène, comme chez *L. lactis* (Hartke *et al.*, 1995) semble indiquer que des mécanismes de réparation de l'ADN sont induits lors de l'ATR. Chez *S. mutans*, différentes études ont révélé qu'en condition d'adaptation à l'acidité : (i) une activité exonucléase de type III codée par le gène *smx* est induite mais son inactivation n'a pas d'effet sur la résistance à l'acidité (Faustoferri *et al.*, 2005) et (ii) la protéine UvrA impliquée dans la réparation des lésions de l'ADN, est induite et son inactivation conduit à une croissance faible à pH 5 et à une incapacité à développer une ATR (Hanna *et al.*, 2001).

Chez les bactéries lactiques, différents mécanismes de résistance à l'acidité ont été mis en évidence. Cependant, les connaissances sur *L. bulgaricus* sont très succinctes et se limitent à des études protéomiques en gel bidimensionnel. La séquence génomique de *L. bulgaricus* sera donc d'une grande utilité pour déterminer quels systèmes de résistance sont conservés avec les autres lactobacilles ou BL.

4. *L. delbrueckii ssp. bulgaricus*

L. delbrueckii ssp. bulgaricus appartient à l'espèce *L. delbrueckii* qui compte trois sous-espèces : *delbrueckii*, *lactis* et *bulgaricus*. *L. delbrueckii ssp. lactis* a le spectre fermentaire le plus large et se retrouve chez les végétaux et les animaux (lait, salive, intestin...). Les sous-espèces *delbrueckii* et *bulgaricus* ont des niches écologiques plus spécialisées : les plantes fermentées pour la sous-espèce *delbrueckii* (par exemple, céréales fermentées ou moût de malt) et le lait pour la sous-espèce *bulgaricus* (de Roissart et Luquet, 1994 ;Tamine et Robinson, 1999).

Dans l'industrie agroalimentaire, *L. bulgaricus* est utilisé pour la fabrication du yaourt en association avec *S. thermophilus*, et dans la fabrication de fromages tels que la mozzarella (Oommen *et al.*, 2002). *L. bulgaricus* est une bactérie homofermentaire stricte, exigeante nutritionnellement, microaérophile et dont la température optimale de croissance est de 42°C. Son pH optimum de croissance se situe autour de 5,5.

Au début de ce travail, peu de données génétiques étaient disponibles sur cette bactérie. Le peu d'outils génétiques utilisables chez *L. bulgaricus* étaient limitant pour les études par génétique classique. La transformation de quelques souches de *L. bulgaricus* est possible mais à des taux relativement faibles, de 10^2 à 10^4 transformants par μg ADN selon les souches (Serror *et al.*, 2002). L'inactivation de gènes est possible (Sasaki *et al.*, 1993) mais nous ne disposons pas de systèmes d'inactivation aléatoire, par exemple par transposition aléatoire.

La législation impose que dans un yaourt, la population de bactéries vivantes soit supérieure à 10^7 bactéries par gramme de produit (source : Codex Alimentarius). La survie des ferments aux différents stress rencontrés dans le processus industriel, y compris lors du stockage du produit fermenté, est donc d'un intérêt économique primordial. En conséquence, différents groupes ont entrepris l'étude des réponses au stress de ce lactobacille. Les données restent cependant descriptives car la plupart des études sont basées sur des analyses globales par électrophorèse bidimensionnelle de protéines. Les protéines dont l'abondance varie durant l'adaptation aux stress n'ont cependant pas pu être identifiées, à l'exception des protéines chaperons (manque de séquences proches et coût du séquençage N-terminal des protéines très élevé) (Silva *et al.*, 2005 ; Gouesbet *et al.*, 2002 ; Lim *et al.*, 2000). Une étude menée au laboratoire a permis d'identifier

une protéine de choc froid (csp) induite lors d'un passage de 42°C à 25°C (Serror *et al.*, 2003). Les réponses au stress de *L. bulgaricus* sont donc encore aujourd'hui largement méconnues.

L. bulgaricus a aussi été étudié pour ses propriétés technologiques. La protéase de paroi PrtB, responsable de la dégradation de la caséine en peptides assimilables par le lactobacille a été bien étudiée. Cette protéase est très importante pour la croissance de *L. bulgaricus* qui est poly-auxotrophe pour les acides aminés (Grobben *et al.*, 1998). Cette protéase a aussi un rôle important pour la croissance de *S. thermophilus* en lait en co-culture avec *L. bulgaricus* (Courtin *et al.*, 2002).

La production d'exopolysaccharides (EPS) est une propriété technologique très importante car ce sont eux qui donne la texture au yaourt. Les EPS sont produits par *S. thermophilus* et par *L. bulgaricus*. Les propriétés et la quantité des EPS produits par *L. bulgaricus* varient entre les souches (Petry *et al.*, 2003). De nombreuses études ont été publiées sur ce sujet et les gènes responsables de la synthèse des EPS ont été caractérisés (Lamothe *et al.*, 2002 ; Petry, communication personnelle).

Le yaourt est à l'origine du concept de probiotique. *L. bulgaricus* a cependant été peu étudié pour ses propriétés probiotiques putatives par rapport à d'autres lactobacilles. Il est connu depuis longtemps que la consommation de yaourt augmente la digestibilité du lactose, notamment chez les personnes intolérantes (Guarner *et al.*, 2005). Cependant, la survie du lactobacille dans le tractus digestif (requis pour une allégation probiotique) reste controversée. Une étude récente semble montrer que *L. bulgaricus* survivrait au transit intestinal (Mater *et al.*, 2005). Certaines souches de *L. bulgaricus* ont une activité immuno-modulatrice qui a été décrite *in vitro* et *in vivo* (souris à flore humaine) mais cette activité semble très variable entre les souches étudiées (Nagafuchi *et al.*, 1999 ; Perdigon *et al.*, 2002).

Mon travail de thèse a comporté deux parties : une partie d'analyses bioinformatiques et une autre, basée sur des études physiologiques. La partie résultats sera donc divisée selon ces deux grandes parties. Dans la première partie, nous aborderons l'annotation du génome et l'analyse des données génomiques, avec notamment un inventaire des gènes présents impliqués dans les réponses au stress. La deuxième partie sera consacrée à l'étude de l'adaptation au stress acide de *L. bulgaricus*.

Résultats – Partie I
Analyse du génome

1. Introduction

Lors de l'initiation de ce travail, peu de données génétiques sur *L. bulgaricus* étaient disponibles et aucune autre séquence génomique complète de lactobacilles n'était encore publiée.

Le génome de la souche ATCC 11842 a fait l'objet d'un projet de séquençage dans le cadre d'une collaboration entre le Centre National de Séquençage et l'INRA. Cette souche a été choisie car elle est considérée comme une souche type de la sous espèce *L. delbrueckii bulgaricus* et elle a été isolée de yaourt (données ATCC). De plus, cette souche est transformable par électroporation à une fréquence de 10^4 transformants/ μg ADN (Serror *et al.*, 2002). Ce projet génomique était important à plus d'un titre, puisque les données génomiques permettent de connaître le répertoire de gènes et de fonctions présents dans la souche, mais aussi puisqu'elles peuvent servir de base pour explorer les variabilités génomiques de l'espèce et permettre d'accéder aux analyses globales via les approches transcriptomique et protéomique.

Ma contribution au projet génomique a d'abord porté sur l'annotation du génome, en parallèle avec les dernières étapes d'assemblage et de finition, réalisées par le CNS et d'autres personnes de l'équipe. Une fois l'annotation réalisée, j'ai recherché plus spécifiquement dans le génome les éléments mettant en évidence une possible adaptation de *L. bulgaricus* à son milieu, le milieu lait. Cette recherche a été axée sur l'étude des capacités métaboliques de la souche ainsi que sur les systèmes de régulation et les mécanismes de résistance aux stress présents.

Cette partie a fait l'objet d'une publication. La partie détaillée ici ne reprend pas les analyses qui sont développées dans cette publication (voir paragraphe 4 p 92).

2. Annotation du génome

2.1. Agmial, présentation et évolution de la plateforme d'annotation

Lorsque nous avons commencé l'annotation du génome de *L. bulgaricus* ATCC11842, nous avons à notre disposition deux systèmes d'annotation : ERGO (Integrated Genomics, <http://ergo.integratedgenomics.com/ERGO/> ; Overbeek *et al.*, 2003) et Agmial, une interface en

cours de développement dans l'unité Mathématique, Informatique et Génome (MIG, INRA). Notre équipe étant une des premières à utiliser le logiciel Agmial, nous avons contribué avec d'autres utilisateurs notamment de l'unité FLEC à son développement en suggérant au fur et à mesure les fonctions à intégrer et qui nous semblaient utiles pour faciliter, compléter et améliorer l'annotation (tableau 5). L'inconvénient est que nous avons été les testeurs de la plateforme, et celle-ci n'a pas toujours été disponible du fait des améliorations successives apportées. Le développement de la plateforme Agmial fait l'objet d'une publication soumise à N.A.R. dont je suis co-auteur qui est présentée en annexe. Pour la clarté de ce manuscrit, je présenterai brièvement Agmial et ses fonctionnalités en mentionnant nos contributions à son évolution. Précisons ici que l'annotation du génome de *L. bulgaricus* ATCC 11842 a été effectuée sur le système Agmial (annotation automatique et manuelle) et a ensuite été communiquée à Integrated Genomics pour transfert dans ERGO.

Tableau 5 : Améliorations apportées à Agmial après discussions entre biologistes et MIG

Fonctionnalités	
	Champs supplémentaires
	Historique annotation
	Champ "note" retranscrit dans le fichier EMBL
	Classe fonctionnelle
	Pseudogène
	Cases à cocher
	Pseudogène
	Gène d'intérêt
	Visible en 2DE
Analyses	
	Composition de la CDS
	% GC, %GC de la 3° base des codons
	CAI (Codon Adaptation Index)
	Homologies
	Recherche de paralogues
	Avec génome proche dont les CDS sont détectées par Agmial
	Dans la base de données CD du NCBI (recherche des COGs)
	Localisation
	Psort
	Reconstructions métaboliques
	Lien direct avec la base de données Pareo

2.1.1. Présentation d'Agmial

Agmial est une plate-forme d'annotation qui regroupe un certain nombre de programmes d'analyse existants ou spécifiquement développés chez MIG. Elle sert donc à la fois à présenter

les résultats de différentes analyses sur la base desquelles une annotation automatique est proposée, et à ajouter ou modifier les données manuellement.

La première étape est la détection des CDS (CoDing Sequences) qui se fait par l'algorithme SHOW, développé par P. Nicolas, et qui est basé sur les chaînes de Markov cachées (<http://genome.jouy.inra.fr/ssb/SHOW/>). SHOW modélise l'alternance de régions codantes et non codantes en associant à chaque CDS détectée une probabilité codante (probabilité que ce soit bien une séquence codante) ainsi qu'une probabilité pour chaque codon initiateur possible de la protéine qui permet d'estimer lequel est le plus probable. Ce logiciel détecte aussi les RBS (Ribosome Binding Sites) sur la base d'un consensus et de la présence d'un codon initiateur d'une CDS en aval. La détection des gènes est ensuite complétée par la détection des terminateurs rho-indépendants par le logiciel Petrin (d'Aubenton Carafa *et al.*, 1990), des ARNt par le logiciel tRNAscan-SE (Lowe et Eddy, 1997) et des ARNr (identifiés par BlastN). Le système propose deux interfaces graphiques, une basée sur Artemis (Rutherford *et al.*, 2000) et l'autre, plus récente, basée sur MuGen (Hoebeker *et al.*, 2003), nous permettant de visualiser chaque gène dans son contexte génomique. La représentation Artemis permet en outre de visualiser le %GC de la région ou le GC skew (figure 10).

Chaque CDS est reliée à une page d'annotation (figure 10) qui contient les données principales nécessaires à l'annotation du gène. Sur cette page, nous avons accès :

- aux propriétés générales de la protéine (nombre d'acides aminés, poids moléculaire, point isoélectrique).
- pour les protéines de plus de 100 acides aminés, aux résultats d'analyses statistiques liées à la CDS comme le %GC, le %GC de la troisième base des codons (GC3%) ou l'index d'adaptation des codons (CAI). Le %GC nous indique si une CDS est biaisée dans sa composition par rapport au reste du génome, ce qui permettrait de détecter les gènes récemment acquis par transferts horizontaux, ceux-ci étant plus riches en A+T (Moszer *et al.*, 1999). Du fait de la dégénérescence du code génétique en particulier sur la 3^e base du codon, celle-ci, comportant moins d'informations codantes, présente en général un biais plus important, indiqué par le GC3%. Le CAI permet de mesurer le biais d'utilisation des codons synonymes à partir d'un groupe de protéines fortement exprimées (en général les protéines ribosomiques). Le CAI de chaque CDS est calculé, et les CDS présentant un CAI élevée sont prédites comme étant fortement exprimées (Sharp et Li, 1987 ; Xia, 1998 ; Carbone *et al.*, 2003).
- aux résultats de recherche d'homologies avec les protéines des banques de données publiques (Swiss-Prot + TrEMBL) sous forme graphique et de tableau.

Protein Summary: 88

Protein References

LB#5#ATCC11842_03.12.2003#4083 in data set Batch 7

Locus Tag

Ldb1201

Protein Annotation

Product:

Gene Name:

Function:

EC Number:

Pseudogene:

Pseudogene number:

Annotation Status: Current Setting is Confirmed

Status:

Comments:

Note:

Comment:

Keywords:

Gene of Interest:

17-12-2002 19:19:35 (stephanie) FC : 2.1.1
 10-02-2003 18:41:02 (stephanie) EC 3.2.1.23.
 10-02-2003 18:41:12 (stephanie) Updated Functional Status
 25-02-2003 17:21:25 (maarten) Updated

General Properties

Length (aa)	1008
CAI using CODON_STAT_CALC(0)	0.0
Clust using CODON_STAT_CALC(0)	1.0
GC% using CODON_STAT_CALC(0)	0.526
GC3% using CODON_STAT_CALC(0)	0.629
GC3dev using CODON_STAT_CALC(0)	-0.0030
Molecular Weight using PI_CALC(0)	114160.35
Isoelectric Point using PI_CALC(0)	5.01

Homology Results (Uniprot) [BLAST_CALC 9]

Display homology results with an expect <=

<input type="checkbox"/>	Beta-galactosidase	ExPASy
<input type="checkbox"/>	LacZ (Fragment). 852 bp	ExPASy
<input type="checkbox"/>	Beta-galactosidase (EC 3.2.1.23). 1044 bp	ExPASy
<input type="checkbox"/>	Beta-galactosidase	ExPASy
<input type="checkbox"/>	Beta-galactosidase (EC 3.2.1.23) (Lactase)	ExPASy
<input type="checkbox"/>	Beta-galactosidase (EC 3.2.1.23) (Lactase)	ExPASy

Identifier	Expect	Identity (%)	Positive (%)	Protein Length	Subject Length	Query Start	Query End	Subject Start	Subject End	Description	Gene Names	Species
Q9X6Y5 ExPASy	0.0	45	62	1008	1023	11	994	7	1012	Beta-galactosidase	LACZ	Bifidobacterium longum
Q6SCN4 ExPASy	0.0	49	66	1008	852	18	834	10	852	LacZ (Fragment). 852 bp	null	null

Homology Results (Lactobacillus gasseri) [BLAST_CALC 5]

Display homology results with an expect <=

Paralogy Results [BLAST_CALC 100]

Display homology results with an expect <=

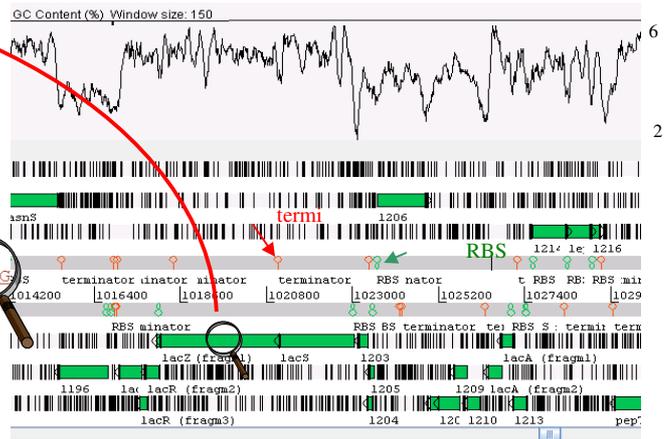
Features Results

- Bgal_small_N using HMMPfam with PFAM
- Glycoside hydrolase, family 42, small chain, N-terminal using InterProScan with Interpro
- Bgal_small_C using HMMPfam with PFAM
- Glycoside hydrolase, family 42, small chain, C-terminal using InterProScan with Interpro
- GLHYDRLASE2 using FPrintScan with PRINTS
- Glycoside hydrolase, family 2 using InterProScan with Interpro
- Glyco_hydro_2 using HMMPfam with PFAM
- Glycosyl hydrolases family 2, immunoglobulin-like beta-sandwich domain using InterProScan with Interpro
- Glyco_hydro_2_C using HMMPfam with PFAM
- Glycosyl hydrolases family 2, TIM barrel domain using InterProScan with Interpro
- Glyco_hydro_2_N using HMMPfam with PFAM
- Glycosyl hydrolases family 2, sugar binding using InterProScan with Interpro
- Transmembrane using MEMSAT

CDD Results [CDD_CALC 0]

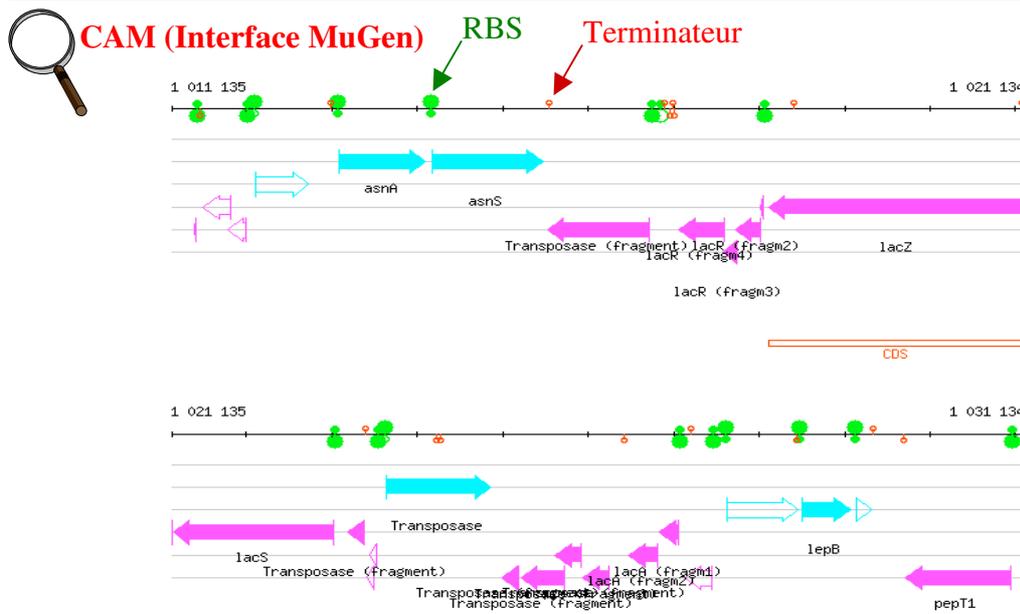
<input type="checkbox"/>	COG3250, LacZ, Beta-galactosidase/beta-glucuronidase [Carbohydrat(...)]	CDD
<input type="checkbox"/>	pfam02836, Glyco_hydro_2_C, Glycosyl hydrolases family 2, TIM (...)	CDD
<input type="checkbox"/>	pfam02837, Glyco_hydro_2_N, Glycosyl hydrolases family 2, sugar (...)	CDD
<input type="checkbox"/>	pfam02929, Bgal_small_N, pfam02929, Bgal_small_N, Beta (...)	CDD
<input type="checkbox"/>	pfam02930, Bgal_small_C, pfam02930, Bgal_small_C, Beta galactosid(...)	CDD
<input type="checkbox"/>	pfam00703, Glyco_hydro_2, Glycosyl hydrolases family 2, (...)	CDD

Interface



Sequence Residues

>LB#88 Beta-galactosidase
 MSNRLVKEKRVQADLAWLTDPEVYEVNITIPPHSDHESFQSQEELEEGKSSLVQSLDGDWLDIDYAENGGG
 PUNFYAEDFDSDNFKSVKVPQNLQGFQGPQYVNVQYVWQSGEEIFPPQIPSKNPLASYVRYFDLDEAF
 WDKVSLKFDGAATAIYVWLNHGFVGYGSDSFTPSEFMVTKFLKKNRRLAVALYKSSASWLEDQDFWR
 MSGFLRVSVTLQAKPRLHLEDLKLASLTDNRYKGLKLEVEANAIYRLPNASFLEVRDSEGLVAEKLGP
 RSEQLFTLADLPAVAASAEKPNLYQVRLYQAGSLLEVSRQEVGFRNFKDGMILYLNQRVIFKGAN
 RHEFDSKLGRAITTEEDMIWDIKMKRSNINAVRCSHYPNQSLFYRLCDKYGLYVIDEANLESHGIWEKVG
 GHEPDSFNVPDGDQHWLGASLSRVKMMARDKNHASIL IWSLGNESYAGTVFAQMDYVRRADPTRVQHY
 EGVTHNRKFDATQIESRMYPAPKVI EYLTNPKAKPFI SVEYAHAMGNSVGDLAAYTALEKYPHYQGGF
 IWDWIDQGLEKDGHLVGGDFDRPTDVEFCGNGLVFADRTESPKNLAVKALYANLKLKLEVDKQGLFLKND
 NLFNTSSSYVFLTSLVDGKLYQSRPLTFGLEPESGTFALPWPEVADEKGEVYRVTAHLKEDLPWAD
 EGFVAAEEVAQKLEPKFPEGRPDVSDYNLGLKGNFQILF SKVQWVPSLVKYAGREYLKRLPEFTF
 WRALTDNRGAGYGYDLARWENAGKYARLKDISEVVKEDSVLVKTAFTLPAVKGDLTVTYEVDRGRKIA
 VTADFPAEEAGLLPAGLNLALPKELTDYRYVYGLGNESYVDRLEGNLYGQAVKRNFSYLPKQET
 GNRSKVRWYQLFDERGQLEFANGADLNLSALPYSAAQIEADHAFELTNNYVWRALSAQMGVGGDDSW
 GQVHPEFCLDAQKRLRLVIQPLLLK



Pareo Enzymes in *Lactobacillus plantarum* *L. bulgaricus*

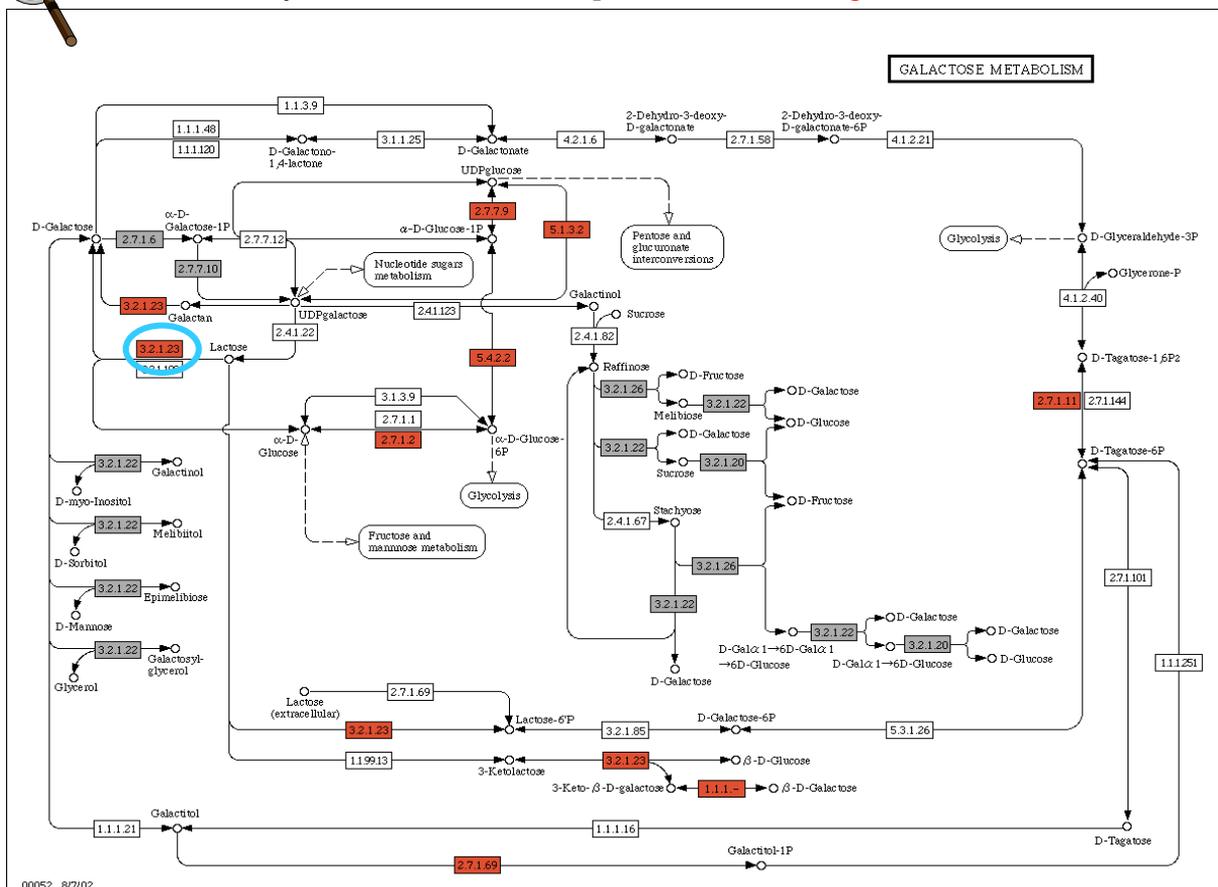


Figure 10 : Interface utilisateur d'Agrial (exemple LacZ)

L'interface artemis permet d'obtenir le contexte génomique et le % en GC de la région (partie haute de la fenêtre, sont indiqués le maximum et le minimum du %GC de la région). Chaque protéine représentée par un rectangle vert est reliée à une page d'annotation.

La page d'annotation (à gauche, partie «Protein Summary») contient les données principales permettant l'annotation.

Le lien CAM nous permet d'avoir accès à une autre représentation (MuGen) du contexte génomique du gène d'intérêt. Le lien vers Pareo nous permet d'avoir accès à la carte KEGG du métabolisme dans lequel est impliquée la protéine (LacZ est entouré en bleu). Dans cette représentation, nous avons la possibilité de comparer les capacités métaboliques de deux organismes, ici *L. plantarum* (enzymes en gris) et *L. bulgaricus* (enzymes en rouge).  Lien cliquable.

- aux résultats de la recherche d'homologie par blastP (Altschul *et al.*, 1997) avec les protéines du génome de *L. bulgaricus* (recherche de paralogues).
- aux domaines identifiés par le programme InterPro Scan (Zdobnov et Apweiler, 2001) ainsi qu'aux domaines conservés avec la base de données CD du NCBI (Marchler-Bauer et Bryant, 2004).

Les protéines potentiellement codées par ces CDS sont ensuite analysées et annotées.

2.1.2. Validation des annotations

L'annotation automatique d'Agmial est basée sur la recherche d'homologie dans les bases de données publiques. Il attribue au champ "product" la description de la protéine qui obtient le meilleur score en analyse blastp (recherche d'homologie entre protéines) et n'est pas désignée comme "hypothetical".

L'annotation manuelle consiste à vérifier, compléter et modifier si besoin l'annotation automatique en vérifiant la cohérence entre les résultats de la recherche d'homologie, de domaines (résultats présentés dans «Feature Result et dans «CDD results», figure 10), le contexte génétique, les paralogues éventuels et les données bibliographiques. L'objectif est que chaque protéine possède à la fin de l'annotation, une identification (champ « product »), et si possible un nom de gène (champ « gene name »), une classe fonctionnelle (champ « fonction ») et un « EC number » dans le cas où le produit est une enzyme (ce champ est nécessaire pour la reconstruction automatique des voies métaboliques). La classification fonctionnelle que nous avons utilisée est basée sur la classification fonctionnelle de SubtiList (<http://genolist.pasteur.fr/SubtiList/>). Cette classification a été légèrement modifiée en collaboration avec l'unité Flore Lactique et Environnement Carné (Annexe 1 p 127). Nous avons rajouté des sous-classes, notamment dans les fonctions de transport, pour avoir une classification plus discriminante dans le cas où les molécules transportées pouvaient être identifiées. Nous avons aussi ajouté certaines classes comme la dégradation protéique, qui, dans le cas de *L. bulgaricus* est importante pour la disponibilité d'acides aminés. Cette classification a aussi été utilisée lors de l'annotation du génome de *L. sakei* (Chaillou *et al.*, 2006). Le contenu du champ «note» d'Agmial, permettant de donner des précisions sur la protéine, a été retranscrit dans le fichier séquence que nous avons soumis à la base de données de l'EMBL (numéro d'accèsion CR954253).

D'autres programmes d'analyse de séquences ont été utilisés au cours de l'annotation manuelle. Nous nous sommes intéressés aux protéines localisées à la membrane ou la paroi et

aux protéines extracellulaires. Ces protéines pourraient jouer un rôle important dans la protocoopération entre *L. bulgaricus* et *S. thermophilus* et/ou dans les contacts directs observés entre les deux bactéries (Bolotin *et al.*, 2004). Toutes les protéines identifiées dans le génome ont été analysées avec le programme SignalP (Bendtsen *et al.*, 2004, <http://www.cbs.dtu.dk/services/SignalP/>) pour la détection des peptides signaux (protéines membranaires ou sécrétées). Cette analyse prédit que 234 protéines (pseudogènes non compris et critère HMM pris en compte) seraient adressées à la membrane. Pour détecter les protéines membranaires, nous avons utilisé les programmes TMHMM (Krogh *et al.*, 2001, <http://www.cbs.dtu.dk/services/TMHMM/>) pour la détection de domaines transmembranaires et PsortB (Gardy *et al.*, 2005) pour la prédiction de la localisation cellulaire. Les résultats de cette analyse sont présentés dans le tableau 9 du paragraphe 4 (p 120). Le croisement des résultats de ces différents programmes nous a permis de prédire une localisation membranaire ou à la paroi pour 56 protéines et une localisation extracellulaire pour 46 protéines.

Pour la validation des annotations, différentes bases de données ont été utilisées. Les numéros EC sont indispensables pour les reconstructions des voies métaboliques et ils évoluent au cours du temps. En effet, certains disparaissent (numéros EC redondants par exemple) et certains sont ajoutés (nouvelle réaction enzymatique identifiée par exemple) et les numéros EC des protéines présentes dans les bases de données publiques ne sont parfois plus valables. Seuls les numéros EC valides figurent sur les cartes permettant les reconstructions métaboliques, la vérification de ces données est donc primordiale pour la confiance accordée ensuite aux reconstructions métaboliques automatiques. 387 protéines du génome de *L. bulgaricus* présentent un EC number (508 si on inclut les pseudogènes). Ces numéros EC ont été systématiquement vérifiés sur la base de données BRENDA, base de données enzymatique de référence dont le contenu est validé manuellement (Schomburg *et al.*, 2002, <http://www.brenda.uni-koeln.de/>). Les annotations de certaines protéines codant des transporteurs ont été validées avec la base de données TCDB (Busch et Saier, 2002, <http://www.tcdb.org/>). Cette base de données regroupe les données sur les transporteurs en les classant en différentes familles et permet, par une analyse par blastP de déterminer à quelle famille appartient la protéine d'intérêt. Pour chaque protéine de *L. bulgaricus* dont l'annotation présentait une incertitude, la famille a été recherchée et la cohérence avec les autres données permettant l'annotation a été vérifiée. Cette analyse a notamment été réalisée pour les systèmes PTS et les systèmes de transport d'hydrates de carbone (20 protéines), ainsi que pour les transporteurs d'ions (32 protéines).

Classiquement, lors des annotations de génome, un seuil pour la taille des CDS détectées est arbitrairement établi. En dessous de ce seuil, les CDS ne sont plus retenues. Ce seuil varie selon les génomes, par exemple, chez *L. plantarum*, il a été établi à 100 nt (33 acides aminés ; Kleerebezem *et al.*, 2003) alors que chez *L. johnsonii*, il est de 225 nt (75 acides aminés ; Pridmore *et al.*, 2004). Le rôle des petits peptides, notamment dans la communication cellulaire et la compétence, est aujourd'hui bien établi (Lyon et Novick, 2004). Pour cette raison, nous avons décidé de ne pas exclure les petites CDS uniquement en fonction de leur taille et nous avons appliqué les critères suivants à chaque protéine de moins de 50 acides aminés : la protéine a été conservée comme produit d'un gène de *L. bulgaricus* si (i) la probabilité codante donnée par SHOW est supérieure à 0,99 et qu'un RBS est présent en amont de l'ATG initiateur ou (ii) si une analyse BlastP ou BlastX permet de détecter une protéine homologue chez une autre bactérie ou (iii) si SHOW prédit une protéine homologue chez *L. gasseri*. La séquence partielle de *L. gasseri* étant disponible sur le site du Joint Genome Institute (la première séquence complète d'un lactobacille du groupe *L. acidophilus*, celle de *L. johnsonii* n'ayant été disponible qu'en février 2004), nous l'avons soumise à la détection de CDS par SHOW. Nous avons observé que les protéines sont souvent très conservées entre *L. bulgaricus* et *L. gasseri* alors que la séquence en ADN ne l'est pas. La présence d'une petite CDS codant une protéine homologue dans le génome de *L. gasseri* permet donc de conforter l'hypothèse de son existence.

Sur 304 CDS de moins de 50 acides aminés, 142 ont été conservées, dont 82 sont associées à des gènes dégénérés et quatre codent des protéines ribosomiques. Les 56 autres sont de fonction inconnue et restent hypothétiques.

Le génome de *L. bulgaricus* présente un nombre important de gènes fragmentés (pseudogènes). Les pseudogènes sont identifiés par homologie de séquence avec les protéines présentes dans les bases de données et sont dus à des mutations entraînant un changement du cadre de lecture, un codon stop précoce ou à des délétions d'une partie du gène. Une dizaine de mutations entraînant des interruptions de cadre de lecture (codon stop ou frameshift) a été vérifiée par séquençage. Toutes les séquences obtenues étaient strictement identiques à la séquence fournie par le CNS. Plusieurs CDS correspondant au même pseudogène peuvent être détectées par le programme SHOW. Ainsi, sur les 2096 CDS détectées, 534 correspondent à des fragments de gènes. Chaque CDS a été annotée, la mention «fragment» étant ajoutée dans le champ «product». Les pseudogènes ont ensuite été reconstitué par blastX et un numéro supplémentaire a été attribué aux CDS les composant, un même numéro correspondant aux différents fragments d'un même pseudogène. 270 pseudogènes ont ainsi été reconstitués. Un

pseudogène étant identifié par rapport à des homologues avec des protéines connues, il est probable que le nombre de pseudogènes dans le génome de *L. bulgaricus* soit sous-estimé. En effet, 288 CDS du génome de *L. bulgaricus* ne présentent aucune homologie avec des protéines des banques de données, il est alors impossible de savoir si ces gènes sont ou non fragmentés.

L'annotation des pseudogènes présente l'avantage de pouvoir prendre en compte ou non les gènes fragmentés, par exemple pour les reconstructions métaboliques. Ne pas prendre en compte les pseudogènes permet d'identifier les capacités de la souche considérée alors que si les pseudogènes sont inclus lors de l'analyse, cela permet de visualiser ce qui peut être possible dans d'autres souches, les gènes inactivés n'étant pas obligatoirement identiques entre les souches. Le génome séquencé est alors considéré comme un exemple des gènes présents dans l'espèce, sans prendre en compte les inactivations récentes de gènes spécifiques à la souche.

Dans le génome, certains gènes semblent présents en plusieurs copies non-identiques montrant une forte homologie entre eux. Une duplication de gène peut être à l'origine de ces copies, qui sont alors appelés paralogues. Ainsi, certains gènes, présents en plusieurs copies dans le génome de *L. bulgaricus*, sont uniques dans d'autres génomes dont les séquences sont disponibles, par exemple les lactobacilles proches. Les paralogues peuvent avoir la même fonction ou s'être spécialisés au cours du temps. Nous avons voulu déterminer quel paralogue était le plus susceptible d'avoir conservé sa fonction de départ. Afin d'annoter correctement les gènes de *L. bulgaricus*, les orthologues (protéines ressemblant le plus aux protéines des autres organismes) ont été identifiés par recherche d'homologie entre la protéine de *L. bulgaricus* et les protéines de quatre organismes de référence arbitrairement choisis (*B. subtilis*, *L. monocytogenes*, *S. pneumoniae*, *C. acetobutylicum*). Dans l'organisme de référence, nous avons sélectionné la protéine qui donne le meilleur score et nous avons cherché son meilleur homologue dans le génome de *L. bulgaricus* (méthode du «reciprocal best hit» Jordan *et al.*, 2002). Si le meilleur score est obtenu pour la protéine de départ, alors les deux gènes sont considérés comme des orthologues et sont susceptibles d'avoir la même fonction. Les autres exemplaires du gène sont considérés comme des paralogues et ont potentiellement une fonction différente.

La recherche d'orthologues a porté sur 31 groupes de paralogues (tableau 6). La souche séquencée présente une répétition inversée parfaite de 23 kbp (voir paragraphe 4). Les gènes compris dans cette répétition n'ont pas été pris en compte dans cette analyse, les protéines codées par chaque répétition étant strictement identiques, elles n'ont été annotées qu'une seule fois. Les gènes identifiés comme orthologues des gènes identifiés chez les autres bactéries

portent le chiffre 1 (par exemple, *acpP1*), les autres exemplaires sont annotés avec les chiffres suivants (par exemple, *acpP2*). Dans trois cas (*pyrR*, *efP* et *folC*), les recherches d'orthologues ne donnent pas de résultats cohérents, c'est-à-dire que l'orthologue dans le génome de *L. bulgaricus* est différent selon le génome de référence considéré. Sur ces trois cas, deux (*pyrR* et *folC*) ont été résolus par analyse du contexte génétique. Les gènes situés dans l'opéron *pyr* (*ldb2108-2114*) et *fol* (*ldb246-249*) respectivement ont été considérés comme les orthologues, les copies isolées dans le génome (*ldb1638* et *ldb0728*) comme les paralogues. Dans le cas de *efP*, les copies ont été numérotées selon leur apparition dans le génome. Nous avons systématiquement regardé si un des paralogues était riche en A/T ce qui pourrait indiquer l'acquisition récente d'une autre copie du gène plutôt qu'une duplication d'un gène déjà présent dans le génome. Une minorité de gènes présente des paralogues riches en A/T. La présence de ce type de gènes n'a pas été considéré comme un critère déterminant puisque les orthologues retenus sont parfois riches en A/T, et notamment ceux identifiés par le contexte génétique.

Tableau 6 : Gènes présentant des paralogues

Gène	Produit	Nb	orthologue identifié	orthologue	paralogue	Gènes riches en A-T
Métabolisme des acides nucléiques						
<i>carA</i>	Carbamoyl-phosphate synthase, small chain	3	oui	<i>ldb2110</i>	<i>ldb1023, ldb1070</i>	<i>ldb2110</i>
<i>pyrR</i>	Uracil phosphoribosyltransferase	2	non	<i>ldb2113</i>	<i>ldb1638</i>	<i>ldb2113</i>
<i>pyrD</i>	Dihydroorotate dehydrogenase	2	oui	<i>ldb2114</i>	<i>ldb1530</i>	<i>ldb2114</i>
<i>udk</i>	Uridine kinase	2	oui	<i>ldb0516</i>	<i>ldb0515</i>	non
<i>gmk</i>	Guanylate kinase	2	oui	<i>ldb1415</i>	<i>ldb2023</i>	non
<i>ndt</i>	nucleoside deoxyribosyltransferase	2	oui	<i>ldb0198</i>	<i>ldb1887</i>	non
<i>prs</i>	Ribose-phosphate pyrophosphokinase	2	oui	<i>ldb0350</i>	<i>ldb0180</i>	non
Enveloppe cellulaire						
<i>acpP</i>	acyl carrier protein	2	oui	<i>ldb1387</i>	<i>ldb0901</i>	non
<i>fabZ</i>	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase	2	oui	<i>ldb0906</i>	<i>ldb1733</i>	non
<i>glf</i>	UDP-galactopyranose mutase	2	oui	<i>ldb1949</i>	<i>ldb2002</i>	non
<i>murE</i>	UDP-N-acetylmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase	2	oui	<i>ldb2044</i>	<i>ldb0699</i>	non
<i>mreB</i>	Cell-shape determining protein	2	oui	<i>ldb0731</i>	<i>ldb0714</i>	non
Métabolisme carboné						
<i>gpmA</i>	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	2	oui	<i>ldb0230</i>	<i>ldb1734</i>	non
<i>gpsA</i>	Glycerol-3-phosphate dehydrogenase (NAD(P)+)	2	oui	<i>ldb0612</i>	<i>ldb1959</i>	non
<i>pfkA</i>	6-phosphofructokinase	2	oui	<i>ldb0838</i>	<i>ldb1772</i>	non
Réponse aux stress						
<i>greA</i>	Transcription elongation factor	2	oui	<i>ldb1485</i>	<i>ldb0290</i>	non
<i>mutS</i>	DNA mismatch repair	2	oui	<i>ldb1616</i>	<i>ldb1603</i>	non
<i>htpX</i>	Protease	2	oui	<i>ldb0160</i>	<i>ldb1185</i>	non
<i>PotA-D</i>	Spermidine/putrescine ABC transporter	2	oui	<i>ldb0647-0650</i>	<i>ldb2178-2181</i>	<i>ldb2178-2181</i>
<i>trxA</i>	thioredoxin	2	oui	<i>ldb1602</i>	<i>ldb1524</i>	non
<i>trxB</i>	thioredoxin reductase	2	oui	<i>ldb0613</i>	<i>ldb1586</i>	non
<i>usp</i>	Universal stress protein	2	oui	<i>ldb0718</i>	<i>ldb0201</i>	non
Divers						
<i>efP</i>	Elongation factor P	2	non	<i>ldb1428</i>	<i>ldb1911</i>	non
<i>rpmG</i>	50S ribosomal protein	2	oui	<i>ldb1482</i>	<i>ldb1669</i>	non
<i>cysK</i>	Cysteine synthase	2	oui	<i>ldb1458</i>	<i>ldb1325</i>	<i>ldb1325</i>
<i>pepT</i>	Tripeptide aminopeptidase	2	oui	<i>ldb1217</i>	<i>ldb0371</i>	non
<i>hsdS</i>	Type I restriction-modification system, specificity subunit	2	oui	<i>ldb1053</i>	<i>ldb1055</i>	<i>ldb1053/ldb1055</i>
<i>sigA</i>	RNA polymerase sigma factor RpoD (Sigma-A)	2	oui	<i>ldb1246</i>	<i>ldb1245</i>	non
<i>crcB</i>	crcB protein involved in chromosome condensation	2	oui	<i>ldb0661</i>	<i>ldb0662</i>	non
<i>parB</i>	Chromosome partitioning protein	2	oui	<i>ldb2053</i>	<i>ldb2055</i>	non
<i>folC</i>	Folylpolyglutamate synthase	2	non	<i>ldb0248</i>	<i>ldb0728</i>	<i>ldb0248</i>

Le nombre de paralogues intacts (Nb) est indiqué. Lorsque la recherche d'orthologue donne des résultats cohérents, l'orthologue est identifié.

Les gènes dupliqués sont répartis sur tout le génome et ne font pas partie de clusters, sauf dans le cas de systèmes dupliqués (systèmes *pot* ou *pyr* par exemple). Les produits des gènes présentant des paralogues sont impliqués dans des fonctions liées au métabolisme des acides nucléiques, à l’enveloppe cellulaire, au métabolisme carboné et aux réponses aux stress.

2.2. Comparaison ERGO/Agmial

2.2.1. Prédiction de gènes

Les deux systèmes détectent les CDS grâce à un algorithme basé sur les chaînes de Markov cachées. Au cours de l’annotation, nous avons pu nous rendre compte que certains gènes détectés par Agmial ne le sont pas par ERGO, sans qu’une comparaison systématique et exhaustive ait été faite. Parmi ces gènes, on trouve des petits gènes, les protéines de choc froid (ldb0647 et ldb2117) par exemple, des gènes chevauchants avec un autre gène, comme *comEB* (ldb0339, gène tardif de compétence), et d’autres gènes comme le gène codant la ribokinase (ldb1697). Il apparaît donc qu’un certain nombre de gènes de fonction connue n’est pas détecté par ERGO.

2.2.2. Annotation automatique

L’annotation automatique d’ERGO est basée sur :

- une recherche d’homologies dans les bases de données d’ERGO (qui incluent les séquences disponibles dans les bases de données publiques et les données génomiques non publiques d’Integrated Genomics),
- la présence de domaines recherchés dans les bases Pfam (Bateman *et al.*, 2002), PROSITE (Falquet *et al.*, 2002), ProDom (Corpet *et al.*, 1998) et COG (Tatusov *et al.*, 2001)
- l’analyse du contexte génétique.

Pour le travail de l’annotateur, un point faible de ce système est qu’aucun commentaire n’explique la raison de l’annotation proposée.

Agmial propose aussi une annotation automatique des gènes qui est toujours basée sur le premier score d’homologie dans les bases de données publiques qui n’est pas annoté comme «hypothetical». Nous avons observé que les annotations d’ERGO et Agmial divergent fréquemment. Du fait que l’annotation fournie par ERGO n’est pas toujours identifiable, nous avons décidé de nous en servir comme information complémentaire pour l’annotation manuelle effectuée dans le système Agmial.

2.2.3. Reconstructions métaboliques

Agmial comporte depuis peu un module permettant les reconstructions métaboliques à partir de la base de données relationnelle Pareo (pathway relational organisation) développée à l'unité MIG de l'INRA, contenant les données métaboliques de KEGG (Kyoto Encyclopedia of Genes and Genomes). A partir des numéros EC assignés à l'étape d'annotation, ce module permet de visualiser les enzymes codées par le génome dans leurs voies métaboliques respectives et de comparer les capacités métaboliques de deux bactéries par coloration différentielle sur la même carte (figure 10).

ERGO intègre un module de reconstructions métaboliques basées comme pour Agmial sur les EC number des enzymes codées dans le génome. Cependant, les gènes fragmentés ou tronqués sont considérés comme présents (et donc actifs), ce qui conduit à une vision des possibilités métaboliques dans d'autres souches mais biaise l'analyse fonctionnelle de la souche au sens strict dont le génome est séquencé.

3. Analyse des données génomiques

3.1. Caractéristiques générales des CDS

L'analyse de l'usage des codons est une analyse classique en génomique. Elle permet d'identifier les biais les plus importants dans l'utilisation des codons. Les résultats de ces analyses diffèrent selon les bactéries, certaines présentant des biais dans l'utilisation des codons importants comme *B. subtilis* (Moszer *et al.*, 1999) ou *E. coli* (Medigue *et al.*, 1991), alors que d'autres n'en présentent pas, comme *H. pylori* (Lafay *et al.*, 2000). Chez *E. coli* et *B. subtilis*, l'analyse en composantes factorielles montrent que trois classes de gènes sont identifiables : les gènes exprimés faiblement, les gènes fortement exprimés et les gènes riches en A-T, prédits comme étant issus de transferts horizontaux. En collaboration avec P. Nicolas (unité MIG, INRA), nous avons réalisé l'analyse en composantes factorielles sur les gènes d'ATCC 11842. Les pseudogènes n'étant probablement plus soumis aux contraintes établissant le biais de codons dans le génome, ils ont été exclus de l'analyse. L'analyse permet d'établir trois classes de gènes distinctes (figure 11A). La dispersion des CDS selon le CAI (Codon Adaptation Index) et le % en GC de la 3^e base permet de discriminer les trois classes (figure 11B), indiquant que ces deux facteurs sont les biais majeurs détectés par l'analyse en composantes factorielles.

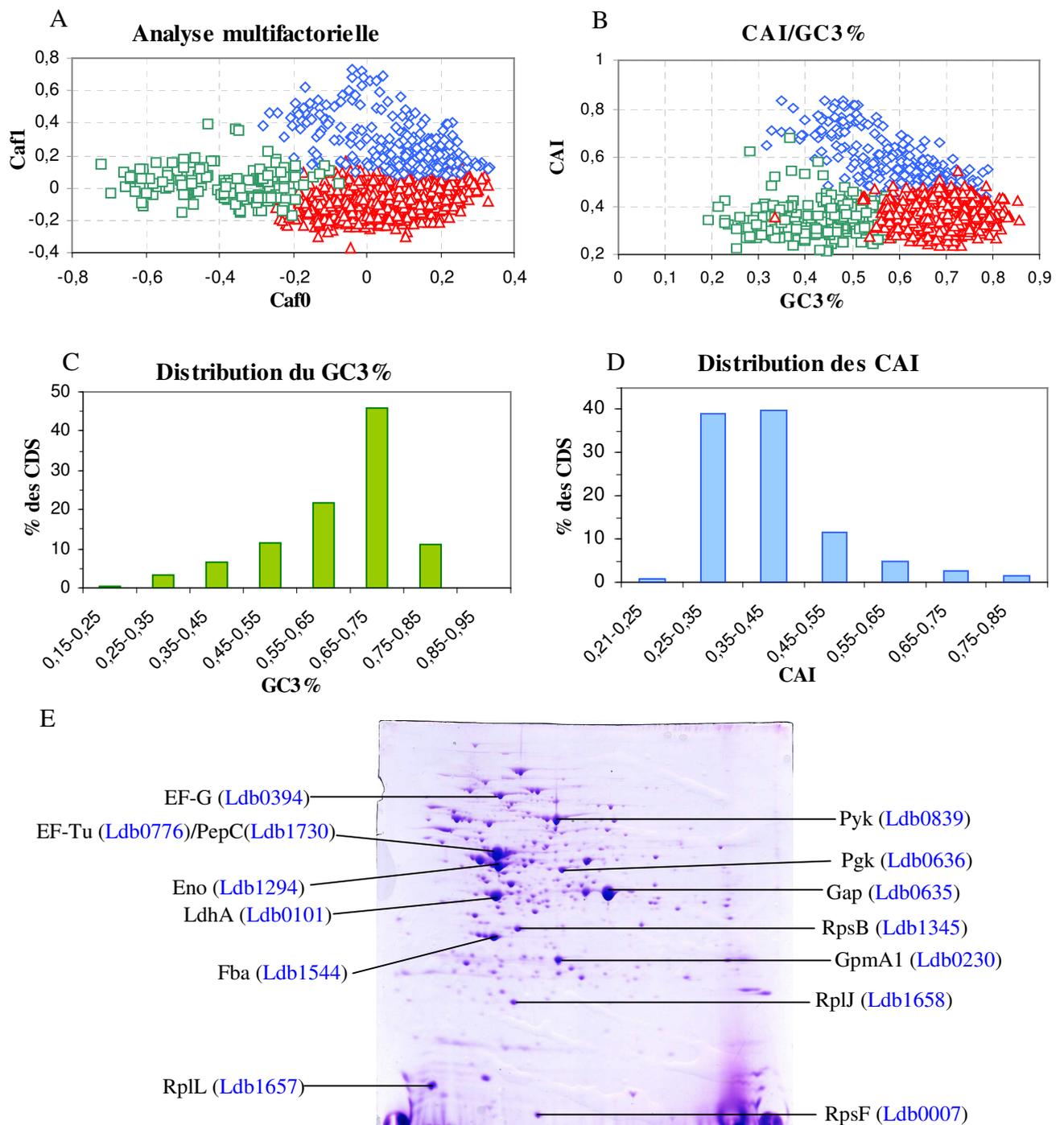


Figure 11 : Analyses de l'usage des codons par *L. bulgaricus*

A-L'analyse de correspondance factorielle permet de définir un plan bi-dimensionnel dans lequel les gènes sont le mieux dispersés (les deux axes Caf1 et Caf0 présentant le maximum d'inertie). Les gènes de la classe I sont en rouge, ceux de la classe II en bleu et ceux de la classe III en vert.

B – La représentation du CAI en fonction du GC3% permet de séparer les trois classes déterminées par l'analyse de correspondance factorielle, indiquant que ces deux facteurs sont les plus représentatifs du biais de codons dans le génome d'ATCC 11842.

C – Distribution des CDS en fonction de leur GC3%.

D – Distribution des CDS en fonction de leur CAI.

E – Exemples de spots abondants caractérisés sur le gel de référence. Les numéros Ldb en bleu signifient que le gène codant la protéine est dans la classe II.

La classe I représente la majorité des gènes de *L. bulgaricus* et est donc considérée comme non biaisée. Nous avons au final considéré comme gènes de classe II et III ceux que l'analyse en composantes factorielles plaçait dans une des deux classes avec une probabilité supérieure à 95% et qui présentent un CAI supérieur à 0,45 pour la classe II (moins de 20% des CDS totales, figure 11D) et un GC3% inférieur à 0,55 pour la classe III (moins de 20% des CDS totales, figure 11C).

La classe II comprend les gènes prédits comme étant fortement exprimés. Elle contient 248 gènes dont ceux codant les protéines ribosomiques, les enzymes de la glycolyse ou les sous-unités de l'ARN polymérase (Annexe 2 p 129). Une analyse protéomique par électrophorèse bidimensionnelle a confirmé que les produits de plusieurs gènes appartenant à cette classe sont très abondants (S. Boudebbouze, com. pers., figure 11E). Les protéines visibles sur le gel de référence n'ont pas encore été toutes identifiées. Néanmoins, 60% des protéines prédites comme appartenant à cette classe et ayant des pI compatibles avec les conditions utilisées ont été identifiées sur gel (Annexe 2 p 129).

La classe III comprend les gènes prédits comme étant issus de transferts horizontaux (Moszer *et al.*, 1999). Elle contient 209 gènes riches en A-T (58,4% par rapport à 49,3% pour l'ensemble des gènes d'ATCC 11842). Les gènes de la classe III sont souvent groupés sur le génome (Annexe 2 p 129) et ces groupes de gènes codent souvent des protéines impliquées dans la même voie métabolique. Citons, à titre d'exemple, la voie de biosynthèse du folate (gènes *ldb0246-0249*), la synthèse des exopolysaccharides (gènes *ldb1998-2008* et *ldb1942-1945*) ou les différentes sous-unités de transporteurs de type ABC (gènes *ldb0251-0254* ou *ldb2178-2181*). Chez ATCC 11842, la classe III comprend des gènes dont nous savons du fait d'autres études réalisées dans l'équipe que leur présence varie suivant les souches. C'est le cas des gènes impliqués dans la biosynthèse des EPS (*ldb1998-2008* ainsi *ldb1942-1945*, S. Petry, com. pers.), les gènes de biosynthèse du folate et d'un ABC transporteur voisin (*ldb0246-0254*, C. Grimaldi, com. pers.), les gènes de biosynthèse des pyrimidines (*ldb2110-2114*, C. Grimaldi et M. van de Guchte, com. pers.) ou les systèmes de restriction-modification (*ldb1053-1055*, C. Grimaldi, com. pers.) et une vingtaine de gènes associés à des IS (C. Grimaldi, com. pers.).

Chez *L. bulgaricus*, l'analyse du biais des codons montre que les biais les plus importants sont le biais traductionnel et une composition atypique des gènes variables entre souches. Cette approche s'est avérée efficace pour prédire les gènes fortement exprimés et les gènes transférés horizontalement.

3.2. Capacités métaboliques

La séquence génomique annotée permet la reconstruction *in silico* des voies métaboliques à partir des numéros EC, ce qui nous donne accès aux capacités métaboliques de la souche. Un nombre important de pseudogènes est impliqué dans le métabolisme carboné (21 pseudogènes) et des acides aminés (14 pseudogènes), nous nous sommes donc intéressés aux capacités de la souche en rapport à ces deux métabolismes. D'autre part, la présence de paralogues dans le métabolisme des acides nucléiques nous a conduit à reconstituer ces voies de biosynthèse. Nous avons donc analysé de plus près les possibilités de la souche pour le métabolisme des sucres (voir paragraphe 4 p 92), le métabolisme des acides aminés et le métabolisme des bases azotées (présentés ci-dessous).

3.2.1. Métabolisme des acides aminés

La reconstruction des voies métaboliques révèle que la souche ATCC 11842 a des capacités de synthèse d'acides aminés très limitées. La souche séquencée ne dispose d'aucun gène nécessaire à la biosynthèse *de novo* de l'alanine, de la glycine, de l'histidine, de l'isoleucine, de la leucine, de la proline ou de la valine. Neuf autres voies de biosynthèse d'acides aminés sont partielles, du fait de la présence de pseudogènes (2 cas) ou de gènes manquants et de pseudogènes (6 cas), ce sont les voies conduisant à l'arginine, la cystéine, la lysine, la méthionine, la phénylalanine, la sérine, le tryptophane et la tyrosine. La souche serait donc incapable de synthétiser ces 16 acides aminés. Les voies de biosynthèse potentiellement fonctionnelles sont celles de l'aspartate, de l'asparagine et de la thréonine (Annexe 3 p 137). ATCC 11842 pourrait aussi synthétiser de la glutamine lorsque du glutamate est présent dans le milieu de culture et inversement.

Ces résultats peuvent être comparés aux résultats obtenus par Grobber et coll. (1998) avec la souche NCFB 2772 de *L. bulgaricus*. Dans cette étude, les exigences nutritionnelles de la souche ont été testées par omission d'un seul composant du milieu chimiquement défini utilisé. Cette souche se développe lorsque le milieu est dépourvu d'aspartate, de glutamate, de glutamine ou de glycine, et dans une moindre mesure d'asparagine ou de thréonine. Seul le résultat avec la glycine est différent des prédictions *in silico* réalisées à partir de la souche ATCC 11842. Il existe deux gènes impliqués dans la biosynthèse de la glycine : un permet sa synthèse à partir de sérine, l'autre à partir de thréonine. Les deux gènes sont absents du génome d'ATCC 11842 mais la présence d'un seul de ces gènes pourrait rendre NCFB 2772 capable de synthétiser la glycine.

3.2.2. Métabolisme des bases azotées

L. bulgaricus semble posséder tous les gènes nécessaires à la synthèse de bases azotées sauf celui codant la nucléoside diphosphate kinase, responsable de la synthèse de dNTP à partir de dNDP. Néanmoins, cette activité pourrait être assurée par l'adenylate kinase comme cela a été montré chez *E. coli* (Willemoes et Kilstrup, 2005). Cependant, les tests de croissance montrent qu'ATCC 11842 ne se développe pas en absence d'uracile (M. van de Guchte, com. pers.), ce qui impliquerait qu'une enzyme dans la voie de biosynthèse de l'UMP à partir de L-Glutamine est inactive.

3.3. Les systèmes de régulation transcriptionnelle

Les systèmes de régulation transcriptionnelle sont très importants pour l'adaptation de la bactérie aux changements de son environnement. L'analyse *in silico* d'un génome peut permettre d'identifier trois types de régulateurs agissant au niveau transcriptionnel : les facteurs σ , les systèmes à deux composants et les régulateurs transcriptionnels. Nous avons recherché ces trois catégories d'éléments dans le génome d'ATCC 11842 ; nos résultats sont présentés dans cette section.

3.3.1. Les facteurs sigma

L'analyse du génome de *L. bulgaricus* ATCC11842, révèle cinq facteurs sigma dont deux facteurs σ primaires Ldb1245 et Ldb1246 dupliqués en tandem, situation unique chez une bactérie à ce jour (voir paragraphe 4). Les lactobacilles séquencés (*L. plantarum*, *L. johnsonii*, *L. acidophilus* et *L. sakei*) semblent présenter peu de facteurs σ (de 1 à 3, Kleerbezem *et al.*, 2003 ; Pridmore *et al.*, 2004 ; Altermann *et al.*, 2005 ; Chaillou *et al.*, 2005) par rapport au microorganisme modèle, *Bacillus subtilis* qui en possède 19 dont 5 impliqués dans le processus de sporulation. Cependant, nous ne pouvons pas totalement exclure que certaines familles de facteurs σ soient encore non identifiées à ce jour.

Ldb1881 est un facteur σ^{ECF} non conservé chez les autres lactobacilles. Il présente une homologie avec σ^{M} de *B. subtilis* (46% de similarité, 25% d'identité). On retrouve des homologues de ce facteur σ chez des bactéries commensales (*Bacteroides thetaiotamicron*) ou pathogènes (*Nocardia farcinica*) de l'Homme. Ce gène présente aussi une faible homologie avec σ^{X} de *Lactococcus lactis* (40% de similarité, 22% d'identité), étudié au laboratoire. La fonction

de ce facteur σ est inconnue. *ldb1881* semble transcrit en opéron avec *ldb1880* qui est homologue à *rsiV*, un anti-facteur σ de *B. subtilis*.

ldb0066 est l'autre facteur σ^{ECF} présent dans le génome de *L. bulgaricus*. On trouve des homologues seulement chez *L. johnsonii* (*lj_1020*, 56% de similarité, 33% d'identité), *L. sakei* (*lsa1509*, 43% de similarité, 22% d'identité) et *L. plantarum* (*lp_1239*, 39% de similarité, 23% d'identité). La fonction de ce facteur σ est inconnue. *Ldb0066* ne semble pas faire partie d'un opéron et ne serait donc pas co-transcrit avec un anti-facteur σ .

Enfin, *ldb1677* est conservé chez *L. acidophilus* (*lba0351*, 51% de similarité, 32% d'identité), *L. sakei* (*lsa1677*, 50% de similarité, 31% d'identité) et *L. plantarum* (*lp_0614*, 49% similarité, 32% d'identité). Le produit du gène *ldb1677* présente des homologies avec σ^{H} de *Listeria monocytogenes* (43% de similarité, 22% d'identité) et de *B. subtilis* (40% de similarité, 21% d'identité).

Les faibles pourcentages de conservation ne nous permettent pas de conclure avec certitude que ce facteur σ est bien orthologue aux autres σ^{H} étudiés chez d'autres bactéries. Nous avons donc examiné le contexte génétique de *ldb1677* et nous l'avons comparé avec celui de ses homologues chez *Staphylococcus aureus* et de *B. subtilis*, ce type de comparaison ayant amené à conclure à l'existence de σ^{H} chez *S. aureus* (Morikawa *et al.*, 2003).

Le contexte génétique du potentiel facteur σ^{H} est conservé entre les trois espèces, ce qui est compatible avec une origine commune de ces régions génomiques et suggère donc que *Ldb1677* appartiendrait bien à la famille des facteurs σ^{H} (figure 12).

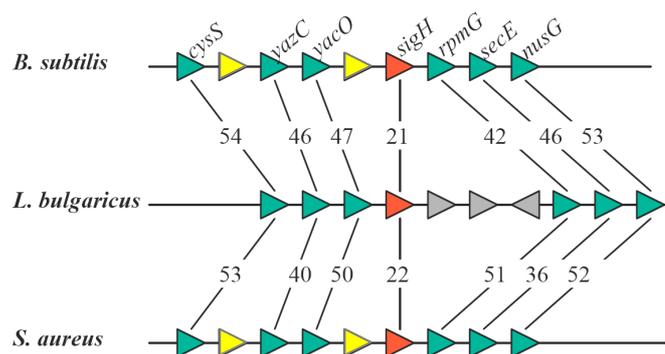


Figure 12 : Conservation du contexte génomique de σ^{H}

Les noms des gènes de *B. subtilis* conservés entre les trois espèces sont indiqués. Les chiffres correspondent au pourcentage d'identité entre les protéines.

Rouge : *sigH* ; Vert : gènes conservés entre les 3 espèces ; Jaune : gènes conservés entre *B. subtilis* et *S. aureus* ; Gris : gènes spécifiques à *L. bulgaricus* codant une protéine hypothétique, une protéine conservée fragmentée et la 1-deoxy-D-xylulose 5-phosphate synthase fragmentée.

Le σ^H de *B. subtilis* est impliqué dans la sporulation, processus n'existant pas chez *L. bulgaricus* ATCC 11842. Le σ^H de *S. aureus* est décrit comme permettant l'induction des gènes tardifs de compétence (Morikawa *et al.*, 2003). Cette fonction est également prise en charge par les ComX de *L. lactis* et *S. pneumoniae* qui appartiennent aussi à la famille σ^H . Bien que *L. bulgaricus* ne soit pas connu pour être naturellement compétent, nous avons recherché dans le génome des gènes de compétence (figure 13 ; Morrison et Lee, 2000).

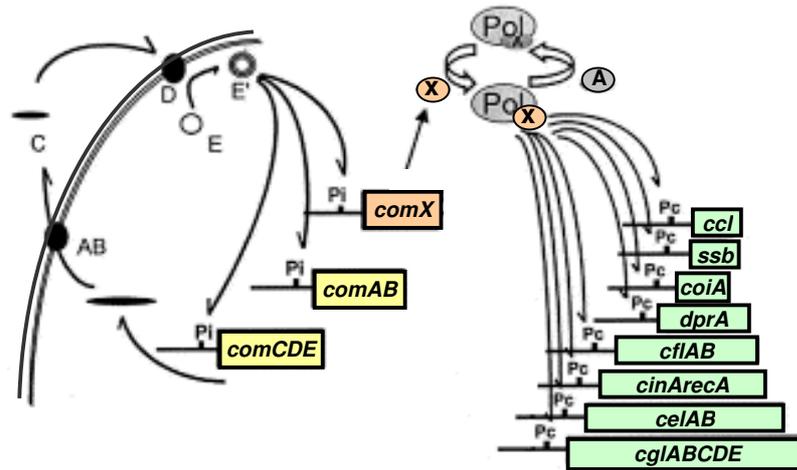


Figure 13 : Régulation des gènes de compétence chez *S. pneumoniae* (d'après Morrison et Lee, 2000).

Les gènes précoces sont colorés en jaune et codent un ABC transporteur (ComAB), une phéromone (ComC) et un système à deux composants (ComDE). L'activation par la phéromone du système à deux composants va induire l'expression de *comX*. Lié à la polymérase, ComX va induire l'expression des gènes tardifs de compétence (colorés en vert).

Pi : Promoteur reconnu par ComE. Pc : Promoteur reconnu par ComX.

La souche ATCC 11842 présente la majorité des gènes tardifs de compétence mais comme chez d'autres BL (*L. lactis* par exemple), les gènes précoces n'ont pas pu être identifiés. Outre *ldb1677* codant le facteur σ putatif, les gènes *smf* (homologue à *dprA*), *coiA*, *ssb*, *recA* et les opérons *comEAC* (homologue à *celAB*), *comFAC* (homologue à *cflAB*), *comGABCD* (homologue à *cglABCD*) sont détectés dans le génome. Les gènes *coiA*, *comFC* et *comGB* sont sous forme de pseudogène et les gènes *comAB*, *comCDE*, *cinA*, *ccl* et *cglE* sont absents. Ces observations suggèrent donc qu'ATCC 11842 ne possède pas un système de compétence naturelle complet. Nous avons néanmoins recherché en amont des gènes tardifs si un motif commun existait en utilisant le logiciel MEME (Bailey et Elkan, 1994 ; <http://meme.sdsc.edu/meme/meme.html>). Le motif TTTTGGCGTANTAGTWAGTG est retrouvé environ 30 pb en amont de l'ATG initiateur des gènes *coiA*, *smf*, et des opérons *comEAC*, *comFAC*, *comGABCD* ainsi que devant *rpsF*, premier gène d'un opéron contenant *ssb*, qui pourrait jouer un rôle dans le devenir de l'ADN

entrant. Ce motif pourrait être le site de liaison du facteur σ Ldb1677. Des études au laboratoire, menées sur la souche IL1403 de *L. lactis*, montrent que la surexpression de ComX induit l'induction des gènes tardifs de compétence ce qui implique que la régulation par ComX est toujours fonctionnelle bien que certains gènes tardifs de compétence soient inactivés dans la souche (Wydau *et al.*, 2006). Il pourrait en être de même chez *L. bulgaricus*.

3.3.2. Les systèmes à deux composants

Le génome de *L. bulgaricus* contient sept systèmes à deux composants, dont deux sont inactivés (un dans l'histidine kinase, l'autre dans le régulateur de réponse). Sur les cinq systèmes à deux composants restant, quatre (Ldb2045/2046, Ldb1492/1493, Ldb0688/0689 et Ldb0135/0136) sont conservés chez les quatre autres lactobacilles séquencés et un (Ldb0878/0879) est conservé seulement chez les lactobacilles du groupe *L. acidophilus*, ainsi que chez *S. pyogenes*. Trois homologues aux systèmes à deux composants d'ATCC 11842 ont été étudiés. Chez *S. agalactiae* et *S. pyogenes*, CsrS/R, homologue de Ldb1492/1493 est impliqué dans la régulation de gènes associés à la virulence mais les régulons ne sont pas conservés entre les deux espèces (Jiang *et al.*, 2005 ;Graham *et al.*, 2002). Chez *S. mutans*, ce système à deux composants est nécessaire à la croissance en condition de stress et son régulon comprend notamment des protéines chaperons (Graham *et al.*, 2002 ; Dalton et Scott, 2004). Chez *L. acidophilus*, l'inactivation de l'histidine kinase de ce système à deux composants conduit à une diminution de la résistance à l'acidité du mutant par rapport à la souche sauvage. L'adaptation de la souche permet néanmoins de restaurer une résistance au stress acide proche de celle de la souche sauvage (Azcarate-Peril *et al.*, 2005). Ce système à deux composants serait impliqué dans la régulation des systèmes protéolytiques et de transport de peptides et d'acides aminés.

YclJ/K de *B. subtilis*, homologue à Ldb0688/0689, est induit en condition d'anaérobiose mais son régulon n'est pas déterminé (Hartig *et al.*, 2004 ; Howell *et al.*, 2003).

YycF/G de *B. subtilis*, homologue à Ldb0135/0136, est un système essentiel à la survie cellulaire impliqué dans l'homéostasie de la paroi, régulant notamment les gènes de biosynthèse des acides téchoïques et des gènes impliqués dans la division cellulaire, *ftsA* et *ftsZ* (Szurmant *et al.*, 2005).

Un des deux systèmes inactivés chez *L. bulgaricus* lui est spécifique (inactivé dans l'histidine kinase), l'autre est homologue au système à deux composants PhoP/R, impliqué dans la régulation de l'opéron de transport à haute affinité du phosphate *pst* chez *B. subtilis* (Qi *et al.*,

1997). Chez ATCC 11842, ce système à deux composants est d'ailleurs situé en aval de l'opéron *pst*.

3.3.3. Les régulateurs

Une comparaison entre les régulateurs présents chez *L. bulgaricus* et *L. plantarum*, premier lactobacille dont la séquence génomique a été disponible est présentée au paragraphe 4. *L. bulgaricus* possède beaucoup moins de régulateurs que *L. plantarum* et certaines familles sont particulièrement sous-représentées. Cependant, *L. plantarum* a un génome beaucoup plus grand que *L. bulgaricus*. Nous avons donc comparé le nombre de régulateurs présents chez *L. acidophilus* et *L. johnsonii*, deux bactéries plus proches de *L. bulgaricus* et ayant un génome du même ordre de taille (1,99 Mb). Ces données n'étant pas présentes de manière complète dans les publications de la séquence de ces deux génomes, nous avons recherché les motifs spécifiques des régulateurs de la base de données Pfam (<http://www.sanger.ac.uk/cgi-bin/Pfam/>) en vérifiant la cohérence avec les données partielles disponibles pour ces génomes et en éliminant les pseudogènes (tableau 7). Les résultats pour *L. johnsonii* sont différents de ceux présentés par Boekhorst et coll. (2004). Les critères pour l'annotation des régulateurs n'étant pas présentés, l'analyse a été entièrement refaite pour être cohérente avec celle des deux autres lactobacilles.

Tableau 7 : Régulateurs transcriptionnels présents dans les lactobacilles du groupe *acidophilus*

Familles	<i>L. bulgaricus</i> (1,86 Mb)	<i>L. johnsonii</i> (1,99 Mb)	<i>L. acidophilus</i> (1,99 Mb)
AraC	2	3	4
ArsR	1	2	2
Crp	3	1	2
DeoR	4	2	3
GntR	4	9	7
LacI	1	7	7
LysR	8	6	7
LytR	4	3	4
MarR	3	6	7
MerR	2	3	3
NagC	0	3	4
Rgg	0	1	1
RpiR	0	5	6
TetR	5	7	6
Xre	3	5	7
Autres	9	5	10
Total	49	68	80

Les fonctions classiquement régulées par ces familles sont disponibles dans le tableau 4 p 35.

Par rapport aux deux autres lactobacilles, *L. bulgaricus* possède significativement moins de régulateurs. Certaines familles de régulateurs sont même absentes du génome de *L. bulgaricus*, comme la famille RpiR ou NagC, d'autres sont sous-représentées, comme LacI, GntR ou MarR. Aucune famille n'est sur-représentée. Les familles de régulateurs RpiR, NagC, LacI et GntR sont généralement impliquées dans la régulation du métabolisme carboné. Ces différences s'expliquent peut-être par les niches écologiques occupées par ces lactobacilles.

L. bulgaricus vit dans le lait, où la source de carbone majoritaire est le lactose. Il possède peu de systèmes de transport et voies cataboliques de sucres, il n'a donc besoin que d'un nombre limité de régulateurs pour ce type de fonctions. Ceci pourrait expliquer le nombre restreint de gènes dans les familles LacI, NagC ou RpiR. *L. johnsonii* et *L. acidophilus* colonisent le tractus digestif humain où les sources de carbone sont diverses, ils possèdent un grand nombre de systèmes de transport et de dégradation de sucres (par exemple, 16 systèmes PTS pour *L. johnsonii* et 20 pour *L. acidophilus* ; Pridmore *et al.*, 2004 ; Altermann *et al.*, 2005) et ont donc besoin d'une régulation fine de ces systèmes. Les régulateurs de la famille MarR sont impliqués dans des mécanismes de résistance à des composés toxiques ou différents stress, notamment par la régulation de mécanismes de transport dont les systèmes d'efflux (Schumacher et Brennan, 2002). *L. bulgaricus* vit dans un milieu plus stable et moins compétitif que les deux autres lactobacilles, la flore digestive étant un écosystème complexe où les mécanismes de résistance aux composés antibactériens pourraient être importants pour la compétitivité.

Certaines familles sont représentées de manière similaire entre les trois lactobacilles, comme les régulateurs de la famille LysR par exemple. Cependant, seul un régulateur de cette famille est commun aux trois espèces, sept sont spécifiques de *L. bulgaricus*, cinq sont spécifiques de *L. acidophilus* et trois sont spécifiques de *L. johnsonii* (annexe 4 p 141). *L. acidophilus* et *L. johnsonii* ont des profils de régulation similaires, le nombre et le type de régulateurs codés par les deux génomes étant relativement proches. Cependant, 35% des régulateurs de *L. acidophilus* et 28% des régulateurs de *L. johnsonii* sont spécifiques de chaque espèce.

3.4. Propriétés probiotiques et survie dans le tractus digestif

L'effet bénéfique du yaourt sur les désordres documenté par Metchnikoff (1907) est à la base du concept de probiotique. Cependant, les activités probiotiques réellement bien établies ou définies restent aujourd'hui rares (Klaenhammer *et al.*, 2005) et les potentielles propriétés probiotiques de *L. bulgaricus* peu documentées et assez controversées et portent essentiellement sur l'amélioration de la tolérance au lactose (Guarner *et al.*, 2005).

Un pré-requis aux propriétés probiotiques d'une bactérie est sa survie dans le tractus digestif. Les études menées pour déterminer si *L. bulgaricus* survivait au passage dans le tractus digestif ont conduit à des résultats contradictoires qui pourraient résulter de différences de méthodes et/ou de variabilité de souches (Mater *et al.*, 2005 ; del Campo *et al.*, 2005 ; Lick *et al.*, 2001). Quelques gènes identifiés par analyse *in silico* dans le génome de *L. bulgaricus* pourraient contribuer à la survie de la bactérie dans le tractus digestif. ATCC 11842 possède une hydrolase de sels biliaires mais elle est sous forme de pseudogène. De plus, il possède deux «multidrug efflux proteines» (*ldb1074* et *ldb1133*), potentiellement impliquées dans l'efflux de composés toxiques comme les sels biliaires. Les autres mécanismes de résistance aux stress rencontrés aussi lors des processus de fabrication seront décrits dans le paragraphe suivant, notamment les systèmes de résistance au stress acide, rencontré en fin de fermentation, pendant le stockage du produit et lors du passage dans l'estomac.

Outre la capacité de survie de la bactérie considérée, nous avons également recherché l'existence de gènes qui pourraient être impliqués dans des activités probiotiques. Il est possible de distinguer trois types de fonctions : (1) celles associées à la surface bactérienne pouvant médier les phénomènes d'adhésion, d'interaction avec l'hôte et d'immunostimulation et (2) la production de métabolites tels que les bactériocines, l'acide lactique, le peroxyde d'hydrogène, des systèmes de quorum-sensing qui peuvent conférer un avantage compétitif dans certains écosystèmes et peut-être avoir des effets anti-bactériens à l'encontre de pathogènes et (3) la synthèse de composés bénéfiques pour l'homme (vitamines, acides gras à courte chaîne...).

En ce qui concerne les fonctions de type (1), nous avons identifié dans le génome d'ATCC 11842 l'homologue de *fbpA* (codant une fibronectin binding protein) de *L. acidophilus*, dont le produit est impliqué dans l'adhésion de cette bactérie aux cellules de type Caco-2 (Buck *et al.*, 2005). L'autre gène impliqué dans l'adhésion de *L. acidophilus* identifié dans cette étude, *mub* (mucin binding), n'a pas été détecté dans le génome. ATCC 11842 possède un opéron *dlt* (*ldb2145* à *2148*) responsable de la synthèse des acides lipotéchoïques contenant de la D-alanine, et qui, chez *L. plantarum*, a été montré comme étant impliqué dans l'immuno-stimulation (Grangette *et al.*, 2005).

En ce qui concerne les fonctions de type (2), nous n'avons identifié aucun gène de biosynthèse de bactériocine. *L. bulgaricus* possède une pyruvate oxydase qui génère du peroxyde d'hydrogène en présence d'oxygène. Nous avons aussi détecté dans le génome le gène *luxS*, dont le produit catalyse la synthèse de l'auto-inducteur 2, responsable du quorum-sensing chez différentes bactéries (pour revue : Xavier et Bassler, 2003 ; Vendeville *et al.*, 2005).

Pour les fonctions de type (3), ATCC 11842 possède un opéron de biosynthèse de folate (vitamine B12), vitamine importante pour l'homme (voir figure 13 p 108). Cependant, nous ne pouvons pas déterminer si *L. bulgaricus* est capable de la synthétiser dans le tractus digestif.

A partir des données génomiques de *L. bulgaricus* ATCC 11842, les éventuelles capacités probiotiques. Cependant, les activités probiotiques sont variables entre souches, comme le montre notamment des études menées sur l'immuno-stimulation par *L. bulgaricus* (Nagafuchi *et al.*, 1999 ; Perdigon *et al.*, 2002). Certaines souches de *L. bulgaricus* pourraient donc posséder des fonctions supplémentaires leur permettant de développer des activités probiotiques.

3.5. Les fonctions impliquées dans les réponses au stress

Au cours des procédés de fabrication (production, stockage et utilisation des ferments, fermentation et stockage du produit), *L. bulgaricus* est probablement soumis à différents stress notamment le stress thermique, le stress oxydant (présence d'oxygène dans le lait), le stress osmotique (changement de milieu des ferments) et le stress acide (acidification du milieu au cours de la croissance). Nous avons recherché dans la séquence génomique de la souche ATCC11842 les mécanismes qui pourraient être mis en œuvre pour résister à ces différents stress.

L. bulgaricus, comme les autres bactéries lactiques étudiées au niveau génomique, ne possède pas d'homologue à σ^B (coordinateur de la réponse générale au stress chez *B. subtilis*) dans son génome. Afin d'identifier des régulons de stress, nous avons à plusieurs reprises au cours de ce travail, recherché des motifs putatifs de régulation chez *L. bulgaricus* et dans les génomes des autres lactobacilles séquencés. Les recherches de motifs de régulation ont été réalisées à l'aide du logiciel iMoMi, développé par N. Pons au sein de l'unité (N. Pons *et al.*, en préparation). Ce logiciel, basé sur MEME (Bailey et Elkan, 1994) permet de rechercher des motifs en amont d'une sélection de gènes mais aussi de rechercher un motif en amont de gènes orthologues chez différents micro-organismes.

3.5.1. Le stress thermique

3.5.1.1. Le stress hyperthermique

Chez les bactéries à Gram +, la réponse au stress hyperthermique fait intervenir deux familles d'acteurs principaux : les protéines chaperons et les complexes multiprotéiques Clp. Chez *L. bulgaricus*, l'analyse *in silico* du génome permet de retrouver ces deux grandes familles

de protéines mais aussi quelques autres protéines référencées comme «Heat Shock Protein» (HSP) chez d'autres micro-organismes. En terme de fonctions potentiellement associées au stress thermique chez *L. bulgaricus*, le bilan est le suivant.

Les protéines chaperons ont pour rôle de permettre aux protéines dénaturées par la température de retrouver leur conformation native. Chez *L. bulgaricus*, les deux principaux opérons de protéines chaperons sont codées par les opérons *groES/groEL* et *hrcA/grpE/dnaK/dnaJ*. Chez de nombreux micro-organismes, *hrcA* code le régulateur négatif de ces deux opérons. Nous reviendrons sur son rôle chez *L. bulgaricus* à la fin de ce paragraphe.

Les autres protéines pouvant être impliquées dans la réponse au stress thermique détectées dans le génome de ATCC 11842 sont :

- La protéine codée par *ldb0289*, homologue à Hsp18, protéine chaperon associée à la membrane, étudiée chez *O. oeni* (Delmas *et al.*, 2001 ; Coucheney *et al.*, 2005). Chez *L. bulgaricus*, *ldb0289* semble être en opéron avec deux gènes codant des peptidases (*pepW* et *pepG*).
- La protéine Ldb0364, homologue à Hsp15 d'*E. coli* chez qui cette protéine est induite en condition de choc thermique (Richmond *et al.*, 1999) et qui s'associe aux ribosomes (Korber *et al.*, 2000).
- La protéine Ldb0369 (codée par *hslO*), homologue à Hsp33, protéine chaperon associée au stress oxydant chez *E. coli* (Winter *et al.*, 2005).

Les complexes multiprotéiques Clp sont la seconde famille de protéines importantes pour la résistance au stress thermique. Les complexes Clp sont composés de deux sous-unités : une sous-unité protéolytique, ClpP et une sous-unité liant l'ATP qui permet la reconnaissance du substrat. Sous forme de complexes, les Clp ont pour fonction de dégrader les protéines qui sont définitivement dénaturées lors du stress. Mais la sous-unité ATPasique peut, sous forme libre, remplir un rôle de protéine chaperon. Citons notamment le cas de ClpB qui ne présente pas le motif (LIV)G(FL) requis pour l'interaction avec ClpP et qui, induite en conditions de stress, aurait un rôle de chaperon protéique (Chastanet *et al.*, 2004). Les gènes *clpP* codant la sous-unité protéolytique et *clpL*, *clpX*, *clpE*, *clpC* codant différentes sous-unités liant l'ATP sont présents dans le génome de *L. bulgaricus*. Les gènes *clpY* et *clpQ*, présents eux aussi, codent pour un système de dégradation indépendant du système ClpP et dont il a récemment été montré qu'il a un rôle mineur dans la résistance aux stress de *S. aureus* (Frees *et al.*, 2005).

Chez *B. subtilis*, bactérie modèle des firmicutes, les régulateurs des protéines chaperons et des complexes Clp sont respectivement HrcA et CtsR. Notre analyse de *L. bulgaricus* révèle la présence d'un homologue de HrcA mais aucun gène susceptible de coder CtsR. Soulignons qu'en amont des *clpP*, *clpC* et *clpE* d'ATCC 11842, aucune séquence comparable au site de fixation de CtsR (^A/_GGTCAAANAN^A/_GGTCAAA, deux dégénérescences admises pour la recherche du motif dans le génome d'ATCC 11842) n'est identifiée. Au contraire, une séquence proche du consensus TTAGCACTC- N₉-GAGTGCTAA correspondant au site de fixation d'HrcA est présente en amont des opérons *groESL* et *hrcA/grpE/dnaK/dnaJ*. Cette séquence est également retrouvée en amont de *clpP*, *clpE* et de *ldb0289* codant l'homologue de Hsp18 (figure 14) mais pas en amont des gènes codant les autres sous-unités Clp (*clpL*, *clpX* et *clpC*), ni en amont de *ldb0364* (Hsp15), de *ldb0369* (*hslO*, Hsp33), *clpY* et *clpQ*. Ces observations nous conduisent à proposer que chez *L. bulgaricus*, le rôle de HrcA soit plus important que chez les autres firmicutes et qu'il pallie en partie l'absence de CtsR. Dans l'état actuel, nous ne pouvons cependant pas affirmer qu'HrcA est le seul régulateur de la réponse au stress thermique chez ATCC 11842. En effet, bien qu'aucune séquence conservée ne soit identifiée en amont de *clpX*, *clpL*, *clpC*, *ldb0364*, *ldb0369*, *clpY* et *clpQ*, il faudrait tester si ces gènes répondent au choc thermique avant de conclure.

<i>clpP</i>	TTAGCACTGTATCTAAATGAGTGCTAA
<i>groES (1)</i>	TTAGCACTTGATCAAGATGAGTGCTAA
<i>groES (2)</i>	TTAGCACTCTCGACTAAAAAGTGCTAT
<i>ldb0289</i>	TTAGCAGTCAAAAAAGTTGAGTGCTAA
<i>clpE</i>	TTAGCACTTGACTAACTAGAGTGCTAA
<i>hrcA</i>	TTAGCATTC TCCTTGACTAACTGCTAA
consensus	TTAGCAcTcnnnnnnnnnnngAgTGCTAa

Figure 14 : séquences CIRCE putatives détectées en amont des gènes codant les protéines de choc thermique

Les bases conservées sont en bleu.

Chez *L. bulgaricus*, la régulation des gènes de réponse au stress hyperthermique semble donc à l'opposé de celle d'*O. oeni* qui ne possède pas HrcA mais dont les opérons codant les chaperons DnaK et GroESL sont sous contrôle de CtsR (Grandvalet *et al.*, 2005).

Nous avons regardé quelle était la situation chez les autres lactobacilles dont les génomes sont séquencés et disponibles. Sur les quatre autres génomes de lactobacilles disponibles, seuls *L. sakei* et *L. plantarum* possèdent CtsR. Les lactobacilles du groupe *L. acidophilus* ne possèdent que HrcA. Chez *L. johnsonii* et *L. acidophilus*, on retrouve des boîtes CIRCE en amont des deux opérons codant les chaperons (dont l'organisation est bien conservée) mais aussi en amont de

clpP et des deux copies de *clpE*. Les homologues de *ldb0289* (*lba0205* et *lj_0181*) semblent eux aussi régulés par HrcA puisqu'ils possèdent une boîte CIRCE. Aucune autre séquence CIRCE n'est détectée dans ces génomes. Le régulon HrcA semble donc conservé au sein du groupe *L. acidophilus*. Chez *L. plantarum* et *L. sakei* qui n'appartiennent pas au groupe *L. acidophilus*, où *ctsR* est présent, on retrouve des boîtes CIRCE en amont des opérons *groESL* et *hrcA-grpE-dnaK-dnaJ* ainsi qu'en amont des homologues de *ldb0289* (*lsa0050*, *lp_3352* et *lp_0129*). Cependant, aucune boîte CIRCE n'est détectée en amont des gènes codant les sous-unités Clp. D'autre part, la boîte de reconnaissance de CtsR est détectée en amont de *ctsR/clpC* et de *clpE* dans ces deux génomes (figure 15).

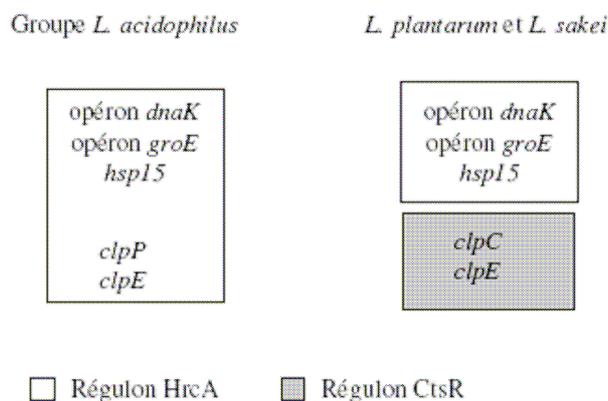


Figure 15 : Modèle de la régulation des opérons de réponse au choc thermique chez les lactobacilles.

Modèle basé sur l'analyse *in silico*.

3.5.1.2. Le stress hypothermique

Outre des modifications du métabolisme cellulaire (Wouters *et al.*, 2000), deux mécanismes de résistance à un stress hypothermique ont été mis en évidence chez les bactéries : le changement de la composition de la membrane en acides gras, et l'induction des «cold shock proteins» (csp).

Lors du choc froid, la proportion d'acides gras insaturés augmente dans la membrane de *B. subtilis* ainsi que les acides gras à courte chaîne carbonée, afin de compenser tout ou partie de la perte de fluidité de la membrane due au froid (Weber *et al.*, 2001). *L. bulgaricus* ne possède pas de désaturase d'acides gras. La séquence ne permet pas de prédire si la synthèse des acides gras peut être modifiée en condition de stress de façon à augmenter la fluidité de la membrane.

Les csp sont de petites protéines qui se lient à l'ARN et rempliraient des fonctions de chaperons ARN, permettant de prévenir l'apparition de structures secondaires dans l'ARN et de

promouvoir l'initiation de la traduction et les phénomènes d'antiterminaison transcriptionnelle (Phadtare, 2004). Le génome de *L. bulgaricus* contient deux gènes codant des *csp*, *cspA* et *cspB* mais seul *cspA* est induit lors de la croissance à basse température (Serror *et al.*, 2003). Le gène *cspA* est homologue aux quatre gènes codant des *csp* identifiés chez *L. plantarum* (trois sont codés par le chromosome et un par un plasmide). Chez *L. plantarum*, seul un des trois gènes *csp* du chromosome est induit en condition de choc froid (Derzelle *et al.*, 2000). *L. acidophilus* et *L. johnsonii* ne présentent, quant à eux, qu'un seul gène *csp*, homologue à *cspB* de *L. bulgaricus*.

3.5.2. Le stress oxydant

Nous avons recherché les protéines décrites comme intervenant dans la résistance au stress oxydant chez d'autres bactéries dans le génome de *L. bulgaricus*. L'ensemble des gènes recherchés dans le génome de *L. bulgaricus* est disponible en annexe 5 p 144.

Les gènes de résistance présents chez ATCC 11842 peuvent être classés suivant les mécanismes de protection impliqués (tableau 9). En effet, la résistance au stress oxydant peut impliquer des mécanismes de détoxification, la synthèse ou le transport de molécules protectrices comme les polyamines, la réduction ou la protection des protéines oxydées.

Tableau 8 : gènes impliqués dans la résistance au stress oxydant présents dans le génome de *L. bulgaricus* ATCC11842

Gène	Fonction	N° de gène	Référence
détoxification			
<i>pox</i>	pyruvate oxidase	<i>ldb2213</i>	De Angelis et Gobbetti, 2004
synthèse de molécules protectrices			
<i>odc</i>	ornithine decarboxylase	<i>ldb1775/ldb0547</i>	Chattopadhyay <i>et al.</i> , 2003
<i>sufB/D</i>	Conserved protein involved in Fe/S cluster assembly	<i>ldb1817/1820</i>	Johnson <i>et al.</i> , 2005
<i>iscU</i>	NifU homolog involved in Fe-S cluster formation	<i>ldb1818</i>	Johnson <i>et al.</i> , 2005
<i>sufS</i>	Cysteine desulfurase / Selenocysteine lyase	<i>ldb1819</i>	Johnson <i>et al.</i> , 2005
<i>sufC</i>	ATP-binding protein involved in Fe/S cluster assembly	<i>ldb1821</i>	Johnson <i>et al.</i> , 2005
transport de molécules protectrices			
<i>pot</i>	spermidine/putrescine ABC transporter ABC-type Mn/Zn transport systems	<i>ldb0647-50/ldb2178-81</i> <i>ldb2037-8</i>	Chattopadhyay <i>et al.</i> , 2003 Ghosal <i>et al.</i> , 2005
<i>flp</i>	transcriptional regulator	<i>ldb0482</i>	Scott <i>et al.</i> , 2000
protection des protéines			
<i>trxB</i>	thioredoxin reductase (2x)	<i>ldb0613/ldb1586</i>	De Angelis et Gobbetti, 2004
<i>trxA</i>	thioredoxin (2x)	<i>ldb1524/ldb1602</i>	Smits <i>et al.</i> , 2005
<i>spx</i>	global regulator	<i>ldb0578</i>	Nakano <i>et al.</i> , 2005
<i>hslO</i>	redox regulated chaperonin	<i>ldb0369</i>	Winter <i>et al.</i> , 2005
<i>msrA</i>	Peptide methionine sulfoxide reductase	<i>ldb0673</i>	Singh <i>et al.</i> , 2003
<i>msrB</i>	Protein-methionine-S-oxide reductase	<i>ldb1971</i>	Singh <i>et al.</i> , 2003
divers			
<i>usp</i>	universal stress protein	<i>ldb0201/ldb0718</i>	Nachin <i>et al.</i> , 2005

L'observation principale est que les gènes codant les enzymes de type peroxydase, superoxide dismutase, catalase ou NADH oxydase, enzymes principales de détoxification de l'oxygène et de ses dérivés toxiques, sont absents du génome de la souche séquencée. Si l'absence de catalase n'est pas surprenante puisque son absence est un critère phénotypique de l'espèce, l'activité NADH oxydase avait été décrite dans la souche B107 de *L. bulgaricus* (Marty-Teyssset *et al.*, 2000). Il est donc étonnant de ne pas identifier le gène correspondant chez ATCC 11842. Comme la NADH oxydase, la pyruvate oxydase permet de transformer l'O₂ en H₂O₂ mais le génome d'ATCC 11842 ne contient pas de gène susceptible d'éliminer l'H₂O₂ ainsi produit. Ces résultats sont cohérents avec des études rapportant la production d'H₂O₂ par les souches de *L. bulgaricus* (van de Guchte *et al.*, 2001 ; Marty-Teyssset *et al.*, 2000), ce qui peut sembler paradoxal puisque *L. bulgaricus* est également décrit comme une espèce assez sensible au H₂O₂ (van de Guchte *et al.*, 2001 ; Marty-Teyssset *et al.*, 2000).

Chez *E. coli*, les polyamines sont importantes pour protéger la cellule des effets toxiques de l'oxygène (Chattopadhyay *et al.*, 2003). L'ornithine décarboxylase est la première enzyme de la voie de biosynthèse des polyamines, elle catalyse la dégradation de l'ornithine en putrescine (qui est une polyamine). ATCC 11842 possède deux gènes codant des ornithine décarboxylases mais elle ne possède pas les gènes nécessaires à la synthèse d'ornithine. L'ornithine pourrait être fournie par *S. thermophilus* qui possède les enzymes nécessaires à sa synthèse. *L. bulgaricus* possède en outre deux systèmes de transport de putrescine et/ou spermidine (*pot*) ce qui semble indiquer que les polyamines jouent un rôle important dans la physiologie de la bactérie. Cette hypothèse est étayée par l'observation qu'un des gènes codant les ornithine décarboxylases (*ldb1775*) est particulièrement biaisé en A-T (35% GC) tout comme un des systèmes *pot* (*ldb2178-2181*) qui ne présente qu'un % GC moyen de 38 % par rapport à 52% pour l'autre système, ce qui pourrait signifier que ces gènes ont été acquis par transfert horizontal. Les ions Zn²⁺ sont impliqués dans la protection contre les effets toxiques de l'oxygène, notamment en protégeant les groupes thiols des protéines (Scott *et al.*, 2000). *L. bulgaricus* possède un ABC transporteur susceptible d'importer les ions Zn²⁺.

Lors d'un stress oxydant, les protéines subissent un phénomène d'oxydation qui peut induire une perte d'activité. Les résidus méthionine notamment, sont susceptibles d'être oxydés en condition de stress oxydant (Singh et Moskovitz, 2003). Deux protéines sont responsables de la réduction des méthionines, MsrA et MsrB. Les gènes codant ces deux protéines sont présents dans le génome d'ATCC 11842. Les thiorédoxines sont impliquées dans la réduction d'un grand nombre d'enzymes et sont elles-mêmes réduites par les thiorédoxines réductases (pour revue : Zeller et Klug, 2006). Chez *E. coli*, la thiorédoxine semble être impliquée dans la résorption des

liaisons di-sulfures induites par un stress oxydant (Ritz *et al.*, 2000). *L. bulgaricus* présente deux thiorédoxines et deux thiorédoxines réductases dans son génome.

Les groupes Fe/S sont impliqués dans l'activité d'un grand nombre d'enzymes en régulant leur potentiel redox ou en étant partie intégrante de leur site catalytique (Johnson *et al.*, 2005). *L. bulgaricus* possède un opéron *suf*, impliqué chez *Erwinia chrysanthemi* dans la biogénèse et la stabilisation des groupes Fe/S lors d'un stress oxydant (Nachin *et al.*, 2003).

En condition de stress oxydant, la protéine chaperon DnaK est inactivée. Chez *E. coli*, la protéine chaperon Hsp33 est activée en conditions oxydantes et compense la perte d'activité de DnaK (Winter *et al.*, 2005). *L. bulgaricus* présente un homologue du gène *hslO* d'*E. coli* codant Hsp33.

UspA (Universal stress protein) d'*E. coli* est impliqué dans la réponse au stress oxydant (Nachin *et al.*, 2005)

Nous avons également recherché dans le génome d'ATCC 11842, la présence de régulateurs souvent associés au contrôle des réponses au stress oxydant. *L. bulgaricus* présente un homologue à *spx* de *B. subtilis* dont le produit, en se liant à l'ARN polymérase, à un rôle de répresseur global. En condition de stress oxydant, il induit spécifiquement certains gènes de résistance dont la thioredoxine et la thioredoxine réductase (Nakano *et al.*, 2005), toutes deux présentes en deux copies dans le génome de *L. bulgaricus*. FlpA est un régulateur de type Crp/Fnr, qui est capable de lier un groupe Fe/S dont l'état d'oxydation régule l'activité du régulateur. Il est impliqué dans la résistance au stress oxydant chez *L. lactis* via le contrôle de l'expression d'une ATPase d'import d'ions Zn^{2+} (Scott *et al.*, 2000). Dans la souche ATCC 11842, on retrouve, en aval de *flp*, un gène codant une ATPase mais le gène semble inactivé par un changement de cadre de lecture. Ce système pourrait cependant être actif dans d'autres souches.

L. bulgaricus tolère la présence d'oxygène puisqu'il est micro-aérophile. Cette caractéristique physiologique semble être concordante avec les données génomiques, qui indiquent que *L. bulgaricus* possède peu de systèmes de détoxification des dérivés oxygénés ce qui le rendrait sensible à ce genre de composés mais qu'il possède différents systèmes de protection contre les effets de ces composés, lui permettant ainsi de continuer à se développer en présence de faible quantité d'oxygène.

3.5.3. Réparation des lésions ADN : le système SOS

A partir de l'étude de Au et coll. (2005) sur le régulon SOS de *B. subtilis*, nous avons recherché les homologues des gènes identifiés dans les génomes des lactobacilles séquencés

ainsi qu'un motif de fixation de LexA, répresseur de la réponse SOS, induite lors des lésions de l'ADN (consensus : GAAC-N₄-GTTC ; Cheo *et al.*, 1991, tableau 9). Contrairement aux lactocoques et aux streptocoques qui ne présentant pas d'homologue de *lexA* identifiable par blast, mais qui ont un analogue appelé *hdiR* (Savijoki *et al.*, 2003), *L. bulgaricus* et les autres lactobacilles possèdent un homologue du gène *lexA*. Les gènes conservés du régulon SOS de *B. subtilis* sont les mêmes pour les cinq lactobacilles (à part pour *yhaZ* qui n'a un homologue que chez *L. sakei*, tableau 9). Selon les lactobacilles, le rôle de LexA pourrait être différent car ses boîtes de fixation potentielles ne sont pas systématiquement retrouvées devant les mêmes gènes.

Tableau 9 : Gènes du régulon SOS de *B. subtilis* conservés chez les lactobacilles

<i>B. subtilis</i>	<i>L. bulgaricus</i>	<i>L. johnsonii</i>	<i>L. acidophilus</i>	<i>L. plantarum</i>	<i>L. sakei</i>
<i>lexA</i>	<i>ldb1356</i>	<i>lj_1508</i>	<i>lba1280</i>	<i>lp_2063</i>	<i>lsa1275</i>
<i>ligA</i>	<i>ldb0466</i>	<i>lj_1720</i>	<i>lba0529</i>	<i>lp_1145</i>	<i>lsa1551</i>
<i>parC</i>	<i>ldb0994</i>	<i>lj_1171</i>	<i>lba1123</i>	<i>lp_1839 (1840)</i>	<i>lsa0975 (0976)</i>
<i>parE</i>	<i>ldb0993</i>	<i>lj_1170</i>	<i>lba1122</i>	<i>lp_1840</i>	<i>lsa0976</i>
<i>pcrA</i>	<i>ldb0465</i>	<i>lj_1721</i>	<i>lba0528</i>	<i>lp_1144</i>	<i>lsa1552</i>
<i>recA</i>	<i>ldb0599</i>	<i>lj_0839</i>	<i>lba0666</i>	<i>lp_2301</i>	<i>lsa0487</i>
<i>ruvA</i>	<i>ldb1614</i>	<i>lj_0467(0465)</i>	<i>lba0410</i>	<i>lp_2287</i>	<i>lsa0366</i>
<i>ruvB</i>	<i>ldb1613</i>	<i>lj_0468(0465)</i>	<i>lba0411</i>	<i>lp_2286</i>	<i>lsa0367</i>
<i>uvrA</i>	<i>ldb0615 (0614)</i>	<i>lj_0864</i>	<i>lba0689</i>	<i>lp_0773</i>	<i>lsa0524</i>
<i>uvrB</i>	<i>ldb0614</i>	<i>lj_0863</i>	<i>lba0688</i>	<i>lp_0772</i>	<i>lsa0523</i>
<i>uvrC</i>	<i>ldb0816</i>	<i>lj_1070</i>	<i>lba0946</i>	<i>lp_2109</i>	<i>lsa1055</i>
<i>yerH</i>	<i>ldb0467</i>	<i>lj_1719</i>	<i>lba0530</i>	<i>lp_1146</i>	<i>lsa1550</i>
<i>yhaM</i>	<i>ldb1538 (f)</i>	<i>lj_1674</i>	<i>lba1589 (1590)</i>	<i>lp_1418</i>	<i>lsa0625</i>
<i>yhaN</i>	<i>ldb1539</i>	<i>lj_1675</i>	<i>lba1590</i>	<i>lp_1417</i>	<i>lsa0624</i>
<i>yhaO</i>	<i>ldb1540</i>	<i>lj_1676</i>	<i>lba1591 (1592)</i>	<i>lp_1416</i>	<i>lsa0623</i>
<i>yhaZ</i>					<i>lsa0809</i>
<i>ynzC</i>	<i>ldb1355</i>	<i>lj_1507</i>	<i>lba1279</i>	<i>lp_2062</i>	<i>lsa1274</i>
<i>yqjW</i>	<i>ldb1610</i>	<i>lj_0471</i>	<i>lba0414 (413)</i>	<i>lp_2280</i>	<i>lsa0382</i>

Les gènes indiqués en bleu sont ceux pour lesquels un motif de fixation de LexA a été identifié. Entre parenthèses apparaît le numéro du gène devant lequel se trouve le motif. (f) : gène fragmenté. N'apparaissent dans le tableau que les gènes de *B. subtilis* qui ont au moins un homologue chez un des lactobacilles considérés (18 sur 33).

3.5.4. Le stress osmotique

Chez les bactéries, la réponse à un environnement présentant une osmolalité élevée passe principalement par l'import ou la synthèse de molécules osmoprotectrices (Sleator et Hill, 2002). Les gènes recherchés dans le génome sont disponibles dans l'annexe 5 p 144.

La réponse la plus rapide à un choc osmotique est l'import massif de K⁺, permettant à la cellule de retrouver sa turgescence initiale (Kempf et Bremer, 1998). Ce phénomène fait intervenir des transporteurs de potassium. Chez *E. coli*, le transporteur majeur de K⁺ en condition de stress osmotique à pH 5,5 est la perméase d'import du K⁺ (Zakharyan et Trchounian, 2001 ; Trchounian et Kobayashi, 1999) dont le gène (*kup*) est présent dans le génome de *L. bulgaricus*

(tableau 10). Les gènes dont les produits sont responsables de l'import de K^+ chez *B. subtilis* (systèmes Ktr ; Holtmann *et al.*, 2003) sont absents du génome de *L. bulgaricus*.

L'import de K^+ est généralement suivi par l'accumulation intracellulaire de composés osmoprotectants tels que la glycine-bétaine (GB), la choline, le glutamate ou la proline (Kempf et Bremer, 1998). Le génome d'ATCC 11842 ne contient pas les gènes permettant de synthétiser la GB, la choline ou la proline mais, comme mentionné précédemment, le glutamate peut être synthétisé à partir de la glutamine. Un transporteur de type ABC, OpuC, susceptible d'importer la glycine-bétaine du milieu extérieur est aussi détecté dans le génome (tableau 10). Soulignons que, d'après la littérature, le lait ne contient pas de glycine-bétaine mais contient son précurseur, la choline (source : <http://www.fao.org/> ; <http://www.regional.org.au/au/asa/2003/p/12/naidu.htm>). Tout comme *S. thermophilus*, *L. bulgaricus* ne possède pas les gènes nécessaires à la transformation de choline en glycine-bétaine. Sur la base de travaux réalisés chez *B. subtilis*, qui montrent qu'OpuC possède une large spécificité de substrat, et notamment qu'il transporte la choline et la carnitine (Kappes *et al.*, 1999), nous pouvons proposer qu'OpuC pourrait également remplir cette fonction chez *L. bulgaricus* qui serait alors susceptible d'utiliser la choline et la carnitine présentes dans le lait comme osmoprotectants. Mentionnons ici que, chez *L. lactis*, le transporteur de type ABC permettant le transport de glutamine transporte aussi du glutamate, qui peut être utilisé comme osmoprotectant. De plus, son activité de transport augmente en condition de stress hyperosmotique (Schuurman-Wolters et Poolman, 2005). *L. bulgaricus* possède deux systèmes de transport de glutamine dans son génome, qui pourraient être impliqués dans le transport de glutamate en réponse à un stress osmotique.

Tableau 10 : Gènes impliqués dans la résistance au stress osmotique présents dans le génome de *L. bulgaricus* ATCC11842

Gène	Fonction	n° de gène	référence
<i>htrA</i>	HtrA-like serine protease	<i>ldb0140</i>	Wonderling <i>et al.</i> , 2004
<i>ftsH</i>	Cell division protein FtsH	<i>ldb0368</i>	Bourdineaud <i>et al.</i> , 2003
<i>opuCAB</i>	Glycine betaine ABC transporter	<i>ldb1180-81</i>	Sleator et Hill, 2002
<i>kup</i>	K+ uptake permease	<i>ldb0219</i>	Trchounian et Kobayashi, 1999
<i>glnPQ</i>	Glutamine ABC transporter	<i>ldb2198-2199</i>	Schuurman-Wolters et Poolman, 2005
<i>glnQHMP</i>	Glutamine ABC transporter	<i>ldb517-521</i>	Schuurman-Wolters et Poolman, 2005
<i>mscL</i>	Large-conductance mechanosensitive channel	<i>ldb0318</i>	Sleator et Hill, 2002
<i>mscS</i>	Mechanosensitive ion channel	<i>ldb1602</i>	Sleator et Hill, 2002

Deux protéases, HtrA et FtsH, pourraient être impliquées dans la réponse au stress osmotique. Chez *L. monocytogenes*, un mutant *htrA*⁻ est plus sensible au stress osmotique (Wonderling *et al.*, 2004). Cette hypersensibilité au stress osmotique serait due à l'accumulation de protéines dénaturées dans le mutant. Chez *B. subtilis* et chez *L. lactis*, un mutant *ftsH* augmente la sensibilité au choc osmotique (Nilsson *et al.*, 1994, Lysenko *et al.*, 1997). Chez *B. subtilis*, des études ont montré que l'induction de *ftsH* était nécessaire à la croissance lors d'un stress osmotique (Deuerling *et al.*, 1995). *L. bulgaricus* possède les gènes codant ces deux protéases.

Lorsque la bactérie se retrouve dans un environnement de basse osmolalité, la turgescence augmente (Sleator et Hill, 2002). Sous l'effet de l'étirement de la membrane, certains canaux (mechanosensitive channels) vont s'ouvrir et permettre l'efflux d'eau et de molécules chargées. *L. bulgaricus* possède deux gènes codant pour ce type de canaux mécano-sensibles, *mscL* et *mscS*.

3.5.5. Le stress acide

Le stress acide est un stress important pour les bactéries lactiques puisqu'elles acidifient leur milieu au cours de la croissance. Les mécanismes mis en jeu pour survivre à l'acidification du milieu ont été décrits dans l'introduction bibliographique ainsi que les gènes identifiés à ce jour comme impliqués dans la tolérance à l'acidité (Annexe 5 p 144). Nous avons recherché ces gènes dans le génome de *L. bulgaricus* (tableau 11).

L'enveloppe cellulaire, interface entre le milieu intra- et extracellulaire, est la plus exposée aux stress environnementaux. Parmi les gènes dont les produits sont impliqués chez les BL dans les modifications de l'enveloppe, *L. bulgaricus* possède le gène *dltC* dont le produit est impliqué dans la formation des acides D-Alanyl-lipotéchoïques de la paroi et dont l'inactivation chez *S. mutans* induit une forte diminution de l'ATR (Boyd *et al.*, 2000a). Les autres gènes étudiés chez les BL, *fabM* et *dagK*, impliqués dans le métabolisme de la membrane, ne sont pas détectés dans le génome d'ATCC 11842.

L'H⁺-ATPase a un rôle majeur dans l'homéostasie du pH cytoplasmique (Kobayashi, 1985). *L. bulgaricus* présente l'opéron codant la H⁺-ATPase (*ldb0705-12*). *L. bulgaricus* possède aussi le gène *ffh*, qui, chez *S. mutans*, est indispensable à une activité normale de l'H⁺-ATPase (Gutierrez *et al.*, 1999). Outre l'H⁺-ATPase, *L. bulgaricus* est particulièrement dépourvu de gènes impliqués dans le maintien de la force protomotrice. Il ne peut pas réaliser de fermentation malolactique ou citrolactique et il ne possède aucune décarboxylase d'acide aminé à part deux ornithines décarboxylases. *L. acidophilus* présente une ornithine décarboxylase dont le rôle dans

la tolérance à l'acidité a été étudiée (Azcarate-Peril *et al.*, 2004). Un mutant de cette décarboxylase est plus sensible à une épreuve acide mais après adaptation, la survie du mutant et de la souche sauvage sont identiques.

De plus, *L. bulgaricus* ne possède aucune voie de production de composés alcalins, les gènes de la voie arginine déiminase et ceux de l'uréase étant absents. Il a été cependant décrit que certaines souches de *S. thermophilus* possèdent une uréase (Mora *et al.*, 2002) et le NH₃ produit par *S. thermophilus* pourrait être transporté par *L. bulgaricus* via son transporteur d'ammonium (codé par le gène *ldb1663*).

Les mécanismes de réparation des dommages causés par l'acidité sont importants pour la tolérance acide. *L. bulgaricus* possède les gènes *recA* et *uvrA*, dont les produits sont impliqués dans la réparation des lésions de l'ADN et les gènes *clpP* et *htrA*, dont les inactivations altèrent la tolérance à l'acidité chez *S. mutans* et qui interviennent dans la dégradation des protéines dénaturées.

Tableau 11 : Gènes impliqués dans la tolérance au stress acide présents dans le génome de *L. bulgaricus* ATCC 11842

Gène	Fonction	n° de gène
Prise en charge des protéines dénaturées		
<i>clpP</i>	ATP-dependent Clp protease proteolytic subunit	<i>ldb0624</i>
<i>htrA</i>	HtrA-like serine protease	<i>ldb0140</i>
Réparation de l'ADN		
<i>recA</i>	Recombinase A	<i>ldb0599</i>
<i>uvrA</i>	Excinuclease ABC, subunit A	<i>ldb0615</i>
Modifications de l'enveloppe cellulaire		
<i>dltC</i>	D-alanyl carrier protein	<i>ldb2146</i>
Maintien de la force protomotrice		
<i>atpA-H</i>	H+ transporting ATPase / ATP synthase	<i>ldb0705-11</i>
<i>odc</i>	Ornithine décarboxylase (2x)	<i>ldb0547/ldb1775</i>
Divers		
<i>ffh</i>	Signal recognition particle protein Ffh	<i>ldb1372</i>

Les gènes considérés sont ceux pour lesquels soit des mutants sensibles au stress acide ont été obtenus soit une inhibition de l'activité du produit du gène conduit à une diminution de la résistance chez d'autres BL.

3.6. Conclusion

Cette analyse des mécanismes de réponses aux stress décrits chez d'autres bactéries et conservés chez *L. bulgaricus* indique que cette bactérie possède relativement peu de systèmes de résistance par rapport aux espèces modèles où ces différents stress sont les plus étudiés. Cette observation est certainement en partie liée à la taille des génomes considérés, *L. bulgaricus* possédant un génome de petite taille (1,86 Mb) comparée à *E. coli* (4,64 Mb pour la souche

K12), bactérie sur laquelle la majorité des études est effectuée, *B. subtilis* (4,21 Mb), bactérie modèle des Gram (+) et même *L. lactis* (2,37 Mb), bactérie modèle des BL.

Autre caractéristique intéressante, *L. bulgaricus* possède un faible nombre de systèmes de régulation. Cette observation pourrait indiquer que les régulons de *L. bulgaricus* seraient plus étendus et/ou moins redondants, dont l'illustration parfaite serait le cas de HrcA, dont le régulon plus étendu engloberait et remplacerait le régulon CtsR présent chez *B. subtilis* ou *L. plantarum*.

Ces observations nous amènent à proposer que *L. bulgaricus*, de part le nombre limité des systèmes mis en jeu au niveau régulation et/ou mécanismes de réponse pourrait être un bon modèle d'étude des réponses aux changements environnementaux.

4. Complete genome sequence of Lactobacillus bulgaricus : evolution caught en route

La publication suivante regroupe les caractéristiques principales du génome de *L. bulgaricus*. Outre les analyses décrites précédemment, ma contribution a porté sur l'analyse des pseudogènes présents dans le génome et sur l'étude des signes de la spécialisation de *L. bulgaricus* au milieu lait.

Cette publication a été acceptée dans PNAS.

Résumé

Lactobacillus delbrueckii ssp. *bulgaricus* (*L. bulgaricus*) appartient au groupe des bactéries produisant de l'acide lactique et est principalement utilisée pour la production de yaourt. La séquence du génome de cette bactérie a été établie et montre des signes d'évolution, avec un nombre important de pseudogènes et de voies métaboliques incomplètes, ainsi que relativement peu de fonctions de régulation.

L'hypothèse que le génome de *L. bulgaricus* subit une phase d'évolution rapide est étayée par différentes caractéristiques du génome : i) Un nombre exceptionnellement important de gènes d'ARNr et d'ARNt par rapport à la taille du génome pourrait indiquer que le génome de *L. bulgaricus* a connu récemment une importante réduction de taille, en accord avec la fréquence importante observée d'inactivation et de perte de gènes, ii) un contenu en GC en position 3 des codons plus élevé qu'attendu sur la base du contenu total en GC suggère que la composition du génome évolue vers un contenu en GC plus élevé, et iii) la présence d'une répétition inversée de 47.5 kbp dans la région de terminaison de la réplication, caractéristique extrêmement rare dans

les génomes bactériens, pourrait être interprétée comme une étape transitoire dans l'évolution du génome.

Les résultats suggèrent une adaptation de *L. bulgaricus*, à partir d'un habitat associé aux plantes, au milieu lait, stable et riche en protéines et en lactose, à travers une perte des fonctions superflues et la protocoopération avec *Streptococcus thermophilus*.

The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution

M. van de Guchte^{*†}, S. Penaud^{*}, C. Grimaldi^{*}, V. Barbe[‡], K. Bryson^{§¶}, P. Nicolas[§], C. Robert[‡], S. Oztas[‡], S. Mangenot[‡], A. Couloux[‡], V. Loux[§], R. Dervyn^{*}, R. Bossy[§], A. Bolotin^{*}, J.-M. Batto^{*}, T. Walunas^{||}, J.-F. Gibrat[§], P. Bessières[§], J. Weissenbach^{†**}, S. D. Ehrlich^{*}, and E. Maguin^{*}

^{*}Génétique Microbienne and [§]Mathématique, Informatique et Génome, Institut National de la Recherche Agronomique, 78352 Jouy en Josas Cedex, France; [‡]Genoscope, 2 Rue Gaston Crémieux, 91057 Evry Cedex, France; ^{||}Integrated Genomics Inc., 2201 West Campbell Park Drive, Chicago, IL 60612; and ^{**}Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8030, 2 Rue Gaston Crémieux, 91057 Evry Cedex, France

Communicated by Todd R. Klaenhammer, North Carolina State University, Raleigh, NC, April 14, 2006 (received for review October 14, 2005)

Lactobacillus delbrueckii ssp. *bulgaricus* (*L. bulgaricus*) is a representative of the group of lactic acid-producing bacteria, mainly known for its worldwide application in yogurt production. The genome sequence of this bacterium has been determined and shows the signs of ongoing specialization, with a substantial number of pseudogenes and incomplete metabolic pathways and relatively few regulatory functions. Several unique features of the *L. bulgaricus* genome support the hypothesis that the genome is in a phase of rapid evolution. (i) Exceptionally high numbers of rRNA and tRNA genes with regard to genome size may indicate that the *L. bulgaricus* genome has known a recent phase of important size reduction, in agreement with the observed high frequency of gene inactivation and elimination; (ii) a much higher GC content at codon position 3 than expected on the basis of the overall GC content suggests that the composition of the genome is evolving toward a higher GC content; and (iii) the presence of a 47.5-kbp inverted repeat in the replication termination region, an extremely rare feature in bacterial genomes, may be interpreted as a transient stage in genome evolution. The results indicate the adaptation of *L. bulgaricus* from a plant-associated habitat to the stable protein and lactose-rich milk environment through the loss of superfluous functions and proto-cooperation with *Streptococcus thermophilus*.

adaptation | GC evolution | large inverted repeat | proto-cooperation | rRNA number

Lactobacillus *delbrueckii* ssp. *bulgaricus* (*L. bulgaricus*) is one of the economically most important representatives of the heterogeneous group of lactic acid bacteria, with a worldwide application in yogurt production. Yogurt has long been recognized as a nutritious, natural, and safe component of a healthy diet and is at the basis of the concept of probiotics (1, 2). A well documented health benefit of the consumption of yogurt containing live *L. bulgaricus* and *Streptococcus thermophilus* is an attenuation of lactose intolerance (3). In addition, immune modulation and diarrhea-alleviating effects have been reported (4), and both *L. bulgaricus* and *S. thermophilus* have been implicated in these effects (3, 5). During yogurt fermentations proto-cooperation between these two bacteria results in an accelerated acidification, but the mechanisms involved are not completely understood (6).

Among the lactic acid bacteria, *L. bulgaricus* belongs to the acidophilus complex, a group of lactobacilli related to *Lactobacillus acidophilus*, *Lactobacillus johnsonii*, and *Lactobacillus gasseri*, which have been used as probiotic cultures. Although within this group *L. bulgaricus* is considered unique because of its atypical GC content, until recently the lack of tools for genetic manipulation has severely hampered a more detailed analysis of this organism (7, 8, ††).

Here we present the genome sequence of *L. bulgaricus* strain ATCC11842, originally isolated from bulgarian yogurt by S. Orla-Jensen in 1919 (unpublished work). The analysis of this genome and comparison to other members of the acidophilus complex and *S.*

thermophilus (9) have contributed to a more complete understanding of its phylogenetic position and revealed important details about its specialized adaptation to milk and proto-cooperation with *S. thermophilus*.

Results and Discussion

Primary Sequence Characteristics. The size of the *L. bulgaricus* genome was estimated at ≈ 1.8 Mbp on the basis of results obtained after pulsed-field gel electrophoresis of chromosomal DNA digests (data not shown). This estimation was confirmed in the present whole-genome sequencing project, where a circular chromosomal sequence of 1,864,998 bp has been assembled. The primary characteristics of this sequence are presented in Table 1.

The overall GC content (49.7%) differs significantly from that of the closely related species *L. acidophilus* (34.7%; ref. 10) and *L. johnsonii* (34.6%; ref. 11), a difference that is mainly due to very important differences at codon position 3 (GC3): 65.0% GC in *L. bulgaricus* as compared with 25.0% and 24.4% in *L. acidophilus* and *L. johnsonii*, respectively. Whereas the latter values fit the strong correlation that can be observed between GC3 and overall GC content in bacteria (Fig. 1), the *L. bulgaricus* GC3 value (65.0%) strongly deviates from the expected value (54.0%). Because the evolution at codon position 3 is generally much faster than at positions 1 and 2, the high GC3 value suggests that the *L. bulgaricus* genome is in an active phase of evolution toward a higher GC content or that in *L. bulgaricus* the correlation between GC3 and overall GC content has been lost or changed.

In the *L. bulgaricus* genome, a large number of coding sequences (CDS) have been annotated as “fragments” on the basis of BLASTP and BLASTX (12) results. These fragments correspond to 270 different pseudogenes (of which 43 are remnants of transposase coding genes) that are regularly distributed over the genome, with a slight underrepresentation in the region from 283 kbp to 672 kbp (Fig. 2). This high number of pseudogenes suggests that the genome is in an active state of gene elimination and concomitant size reduction. Not counting pseudogenes, only 73% of the genome consists of (putative) coding sequences. A distinct 2.5-kbp noncoding region (starting at position 764 kbp) was identified that has all of the features of a CRISPR region (clustered regularly interspaced

Conflict of interest statement: No conflicts declared.

Abbreviations: CDS, coding sequences; IS, insertion sequence.

Data deposition: The *L. bulgaricus* genome sequence has been submitted to the European Molecular Biology Laboratory database (accession no. CR954253).

[†]To whom correspondence should be addressed. E-mail: maarten.vandeguchte@jouy.inra.fr.

[¶]Present address: Department of Computer Science, University College London, London WC1E 6BT, United Kingdom.

^{††}Sasaki, T., Ito, Y., & Sasaki, Y. (1993) *FEMS Microbiol. Rev.* 12, P8.

© 2006 by The National Academy of Sciences of the USA

Table 1. Characteristics of the *L. bulgaricus* ATCC11842 genome

Genome size, bp	1,864,998
Overall GC content, %	49.7
GC content of CDS, * %	51.6
GC content of CDS* at codon position 3, %	65.0†
Number of CDS*	1562
CDS* as % of genome sequence	73
Number of CDS* with unknown function	598
Number of pseudogenes‡	270
Number of <i>rrn</i> operons	9
Number of tRNA genes	95

*CDS not annotated as “fragment.”

†Corresponding to 534 CDS annotated as “fragment.”

‡64.4% if pseudogenes are included in the analysis.

short palindromic repeats) (13). CRISPRs have been interpreted as traces of past invasions by extrachromosomal elements and have been hypothesized to provide immunity against foreign DNA expression by coding antisense RNA (14).

L. bulgaricus contains a relatively high number of rRNA and tRNA genes. In the firmicutes, a clear correlation can be observed between the numbers of rRNA and tRNA genes and between each of these and genome size (Fig. 3 and Fig. 5, which is published as supporting information on the PNAS web site). Although in *L. bulgaricus* the first correlation is respected, the numbers of tRNA and rRNA genes are ≈50% higher than the average and 20–30% higher than the highest values observed so far for this genome size, corresponding to values found for genomes of 3–4 Mbp in size. Whereas variation in rRNA and tRNA gene copy numbers may be related to the capacity to respond to changing environmental conditions (15, 16), the exceptionally high copy numbers found in *L. bulgaricus* likely indicate that the genome has undergone a recent phase of size reduction.

Genome Structure and Organization: Replication Terminus with a 47.5-kbp Inverted Repeat and Duplicated *dif* Sites. A sharp change in the sign of the GC skew coincides with the location of the *dnaA* and *dnaN* genes (data not shown), indicating the likely presence of the origin of replication in this region. The presence of several DnaA boxes in this region confirms this hypothesis (17). On the opposite side of the genome, a second change in the sign of the GC skew locates the replication terminus in the region between positions 940000 and 946000 (data not shown). An intriguing feature of the replication terminus region of the ATCC11842 genome is that it

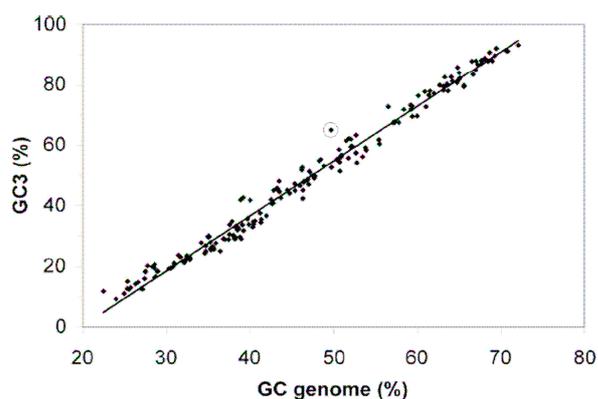


Fig. 1. Relationship between GC content at position 3 of coding sequences (GC3) and genomic GC content in 232 eubacterial genomes. The *L. bulgaricus* value was calculated excluding pseudogenes. The *L. bulgaricus* data point is circled.

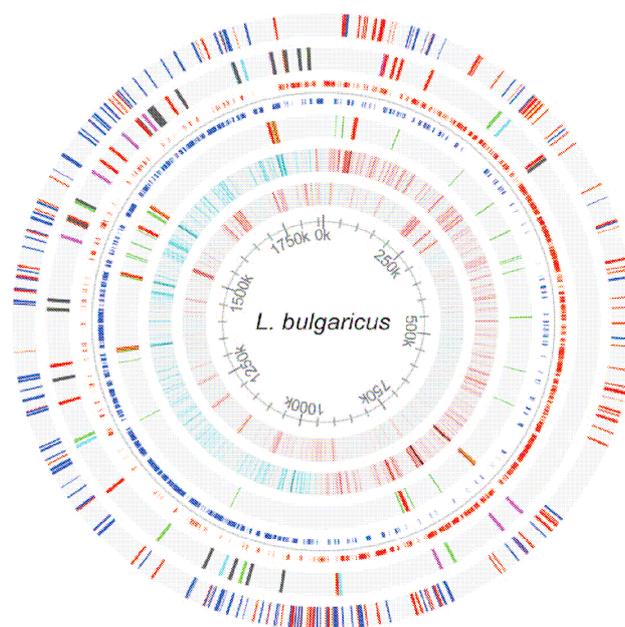


Fig. 2. Genome atlas of the *L. bulgaricus* genome. The seven circles (outer to inner) show the following. Circle 1, pseudogenes on positive (red) or negative (blue) strand. Circle 2, IS elements (transposases or ISL4-related hypothetical genes). Elements with fewer than four copies are represented in gray, and elements with more than five copies are represented by separate colors of red (ISL7), purple (ISL4), blue (ISL5), and green (ISL4–5). (See also Table 2.) Circle 3, CDS (excluding pseudogenes and transposases) on positive (red) or negative (blue) strand. Circle 4, rRNA (red) and tRNA (green) genes. Circle 5, [(G-C)/window size (2000)], from less than -0.1 (cyan) to more than $+0.1$ (red). Circle 6, [(A+T)/window size (500)], from <0.3 (cyan) to >0.7 (red). Circle 7, position on the genome. The genome atlas was constructed by using GENEWIZ software (53).

contains an inverted repeat of 47.5 kbp (positions 918952–966484) enclosing a unique central region of 1.4 kbp between the repeated sequences (23 kbp each; see Fig. 6, which is published as supporting information on the PNAS web site). In bacterial genomes, inverted repeats of this size are extremely rare. To our knowledge, they have been observed only in the replication terminus region of certain *Streptomyces* chromosomes that have circularized after telomere deletion (18). Preliminary results of one primer PCR amplification in 30 different *L. bulgaricus* strains show that an inverted repeat is conserved in most or all strains (data not shown). The size of the unique central sequence and/or the size of the repeat may largely vary between strains, however, and this feature may represent a transient stage in the evolution of the *L. bulgaricus* genome.

In *L. bulgaricus* the repeated sequence includes the putative *dif* site (Figs. 6 and 7, which are published as supporting information on the PNAS web site), which is the recombination site involved in the resolution of chromosome dimers (19). Consequently, *L. bulgaricus* contains two putative *dif* sites in opposite orientation at a distance of 2.5 kbp (0.55 kbp of repeated sequence at either side of the unique central sequence of 1.4 kbp) instead of the unique site found in other microbial genomes, such as *Bacillus subtilis* (Fig. 7). One may hypothesize that this configuration could interfere with the efficient segregation of dimeric chromosomes that may be formed during replication, because recombination between two *dif* sites of opposite orientation in a chromosome dimer would result in the inversion of a part of the dimer rather than resolution of the dimer. Such FtsK XerCD-dependent inversions have been observed in plasmids containing *dif* sites of opposite orientation *in vitro* (20). The results of PCR analysis (data not shown) suggest that

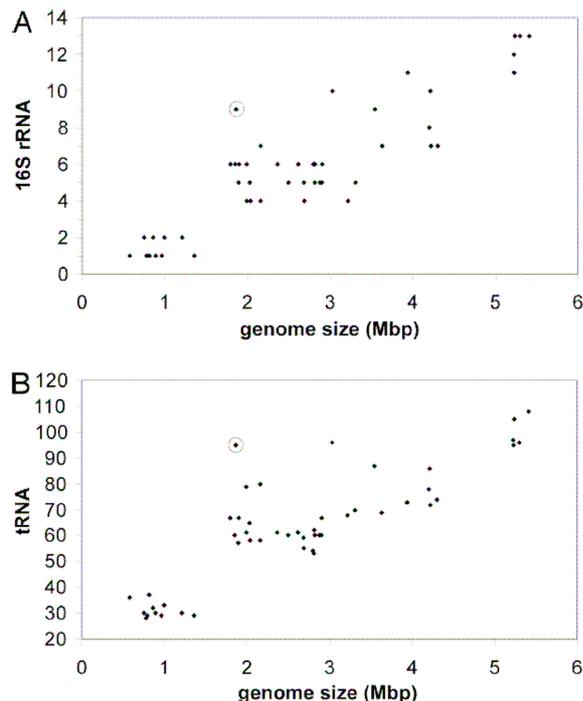


Fig. 3. rRNA and tRNA genes as a function of genome size. (A) Relationship between the number of 16S rRNA genes and genome size in 54 firmicutes genomes. (B) Relationship between the number of tRNA genes and genome size in 54 firmicutes genomes. *L. bulgaricus* (1.9 Mbp, 9 16S rRNA, and 95 tRNA) data points are circled. For comparison: *L. acidophilus* (2.0 Mbp, 4 16S rRNA, and 61 tRNA), *L. johnsonii* (2.0 Mbp, 6 16S rRNA, and 79 tRNA), and *L. plantarum* (3.3 Mbp, 5 16S rRNA, and 70 tRNA).

the 1.4-kbp unique sequence enclosed by the repeated sequences may be present in either of two orientations in a population of strain ATCC11842 (Fig. 6). This result would be expected as a consequence of such a site-specific recombination between two *dif* sites of opposite orientation or, alternatively, of homologous recombination between the inverted repeats.

Nonrandom Distribution of Mobile Elements. The *L. bulgaricus* ATCC11842 genome contains a large number of transposases and remnants thereof (Table 2, which is published as supporting information on the PNAS web site), most of which have been described earlier in *Lactobacillus delbrueckii* insertion sequence (IS) elements (ISL3, ISL4, ISL5, ISL4–5, ISL6, ISL7, and ISld11) (21–23). In the present work, seemingly intact transposases of two new types could be detected [named ISL8 (Ldb0326 and Ldb2033) and ISL9 (Ldb1987)], as well as remnants of several other types.

Intriguingly, a region of 415 kbp (22% of the genome) between positions 279 kbp and 694 kbp is completely exempt of IS elements, whereas (remnants of) the more abundant IS elements seem randomly distributed over the remaining 78% of the genome (Fig. 2). This discontinuity in the distribution of IS elements coincides with the region, mentioned above, where pseudogenes are less abundant. The absence of IS elements does not, however, explain the difference in the frequency of pseudogenes, because in the whole genome inactivation of only 9 of 227 nontransposase pseudogenes occurred from insertion of IS elements. The positions of transposases often coincide with regions of locally elevated A+T content (Fig. 2 and Fig. 8, which is published as supporting information on the PNAS web site), often considered as possible signs of horizontal gene transfer.

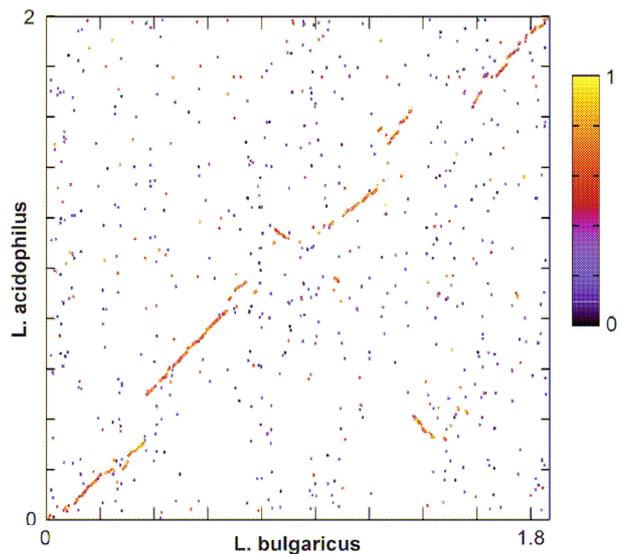


Fig. 4. Synteny between *L. bulgaricus* and *L. acidophilus* genomes. x axis, position on *L. bulgaricus* genome (Mbp); y axis, position on *L. acidophilus* genome (Mbp). 0, replication origin. Colors indicate protein similarity by BLAST score ratio (25) according to the scale on the right.

In contrast to many other lactic acid bacteria, *L. bulgaricus* ATCC11842 does not contain any prophage. Only four pseudogenes of putative phage origin were detected in a large region around the replication terminus, where the *attB* sites of the elsewhere-described bacteriophages mv4 (24) and JCL1032 (K.-A. Riipinen, personal communication) are found.

Synteny Among *L. bulgaricus*, *L. acidophilus*, and *L. johnsonii* Genomes.

A clear global synteny is observed among the similarly sized genomes of *L. bulgaricus* and *L. acidophilus* or *L. johnsonii* (Fig. 4 and Fig. 9, which is published as supporting information on the PNAS web site). The most important perturbations of synteny are found in a region of \approx 300 kbp around the replication terminus of the *L. bulgaricus* genome and in two smaller regions at \approx 400 kbp at either side of the replication origin. By using the criterion of BLAST score ratio with a cutoff value of 0.4 (25) or the results of a BLASTP reciprocal best-hit analysis, \approx 55–60% of the *L. bulgaricus* proteins can be considered as having homologues in *L. acidophilus* and *L. johnsonii* and constitute the common backbone of these three closely related genomes. Between 25% and 35% of the proteins appear unique to *L. bulgaricus* (Fig. 10 and Tables 3–5, which are published as supporting information on the PNAS web site). Although the differences between these genomes are likely to be of importance in their respective ecological niches, they mainly concern unknown functions. The most marked differences among known functions comprise the presence of complete pathways for the biosynthesis of folate and saturated fatty acids (see below) in *L. bulgaricus*, both of which are partially lacking in *L. acidophilus* and *L. johnsonii*. Likewise, *L. bulgaricus* contains all or most of the enzymes necessary for the biosynthesis of purines and pyrimidines, respectively, of which several are absent from *L. acidophilus* and *L. johnsonii* and can grow in a purine-free medium (data not shown). *L. bulgaricus* contains *ppx* and *ppk* genes that are involved in the turnover of polyphosphate and thereby may play a role in adaptation to stress conditions (26). Inversely, proteins of known function that are present in *L. acidophilus* or *L. johnsonii* but not in *L. bulgaricus* are mainly involved in sugar transport and metabolism (data not shown).

Restriction Modification Systems. The *L. bulgaricus* genome contains a complete type I restriction modification (R/M) system (Ldb1051, Ldb1052, and Ldb1053), an additional type I specificity subunit (Ldb1055), and a putative Mrr type restriction endonuclease (Ldb0474). The presence of these R/M systems may explain part of the difficulties encountered previously to transform this strain. A type III restriction modification system appears inactive as a consequence of mutations in the gene encoding the endonuclease (Ldb1227–Ldb1229), whereas the corresponding methyltransferase (Ldb1230) seems intact.

Regulatory Functions: Two SigA Homologues and Relatively Few Regulators. A unique feature of the *L. bulgaricus* genome is that it encodes two 58% identical SigA homologues in two genes that are organized in tandem and apparently cotranscribed. While one of the two encoded proteins (Ldb1246) shows –35 and –10 recognition domains that are 100% identical to those in *B. subtilis* SigA, the other (Ldb1245) shows some important differences in these domains (Fig. 11, which is published as supporting information on the PNAS web site). It is unknown whether this second copy is active and recognizes different promoters. If so, it may be considered a new σ factor developed after gene duplication. Apart from these SigA homologues, two putative extracytoplasmic function (ECF)-type σ factors (Ldb0066 and Ldb1881), one putative σ factor of unidentified type (Ldb1677), and one anti- σ factor (Ldb1880) were found.

Relatively few genes encoding transcriptional regulators were identified in *L. bulgaricus*, and the difference with *Lactobacillus plantarum* (27) was especially striking (53 and 234 predicted genes, respectively; Table 6, which is published as supporting information on the PNAS web site), even when considering the smaller genome size of *L. bulgaricus* (1.9 Mb as opposed to 3.3 Mb for *L. plantarum*). Both these features likely reflect adaptations to the stable and nutritionally rich milk environment, where fewer biosynthetic functions and less adaptive regulation are required. The largely different sugar metabolic capacities of these two species are reflected by the difference in the number of LacI type regulators. Notable differences in MarR-, MerR-, TetR-, and ArsR-type regulators may indicate significantly different levels of resistance to a variety of compounds and stress conditions or differing strategies to cope with them. Regulatory circuits may have a less complicated structure in *L. bulgaricus*, as exemplified by the predicted HrcA regulon (Table 7, which is published as supporting information on the PNAS web site) (28). In *B. subtilis*, *dnaK* and *groESL* on one hand and *clp* on the other hand are under the control of two different regulators, HrcA and CtsR, respectively. *L. bulgaricus* encodes only a HrcA homologue. CIRCE boxes, the operators for HrcA (29), are found upstream of *dnaK* and *groESL* and upstream of *clpP* and *clpE*, suggesting a broader role for HrcA in *L. bulgaricus*. A similar organization was found in *L. acidophilus* and *L. johnsonii* (Table 7) (10, 11).

Pseudogenes and Specialization Through Loss of Function. *L. bulgaricus* strain ATCC11842 was originally isolated from Bulgarian yogurt. Traditional yogurt-making involves the sequential transfer of samples of yogurt cultures to fresh milk. With the first records of yogurt (kisim) dating to 3200 before Christ (30), one would logically predict that *L. bulgaricus* adapted to this environment over time. Traces of such an adaptation process were found throughout the *L. bulgaricus* genome in the form of pseudogenes and incomplete metabolic pathways.

A total of 270 pseudogenes have been detected in the *L. bulgaricus* genome, including 43 transposases. The remaining 227 genes represent 12% of the total number of protein-coding genes. This percentage of pseudogenes is remarkably high (31) and may indicate a recent and ongoing process of specialization. A function, albeit inactivated, could be assigned to 81 pseudogenes. For 30 of these genes, intact paralogues are present in the genome whereas

for the remaining 51 genes no intact paralogues could be detected, indicating that the corresponding functions are completely lost (Table 8, which is published as supporting information on the PNAS web site). These include genes implicated in carbohydrate metabolism, amino acid and cofactor biosynthesis, and competence development. For 146 pseudogenes it was difficult to determine whether their function could be assumed by a paralogue as only a general function (e.g., “ABC transporter”) or no function at all could be attributed to these genes.

Complete transport systems for the milk sugar lactose (lactose permease), mannose/glucose [phosphotransferase system (PTS) IIABCD], fructose (PTS IIABC), and glycerol (glycerol uptake facilitator) are present, but several other transport systems appeared incomplete. Remnants were found of cellobiose (PTS subunit IIC), sucrose (PTS subunit IIA), maltose (partial subunits of a dedicated ABC transporter), and some other sugar transport systems of undetermined specificity (several isolated PTS IIA subunits). Together with remnants (pseudogenes) of key enzymes in carbohydrate-specific metabolic pathways (glycerol kinase, 6-phospho- β -glucosidase/galactosidase, α -amylase, and mannitol-1-phosphate 5-dehydrogenase), these reveal the prior existence of metabolic capacities that may notably have served in a plant-associated environment.

Apart from the intrinsic evidence for specialization through loss of function provided by the presence of pseudogenes, the complete absence of a large number of enzymes involved in the biosynthesis of amino acids (incomplete or lacking pathways) also suggests an adaptation to the protein-rich milk environment. Here the presence of an extracellular protease, several amino acid transport systems, two complete peptide transport systems, and numerous peptidases (Tables 9 and 10, which are published as supporting information on the PNAS web site) would render most amino acid biosynthesis pathways superfluous. As a consequence, the extracellular protease has become essential for growth in milk (32).

Interestingly, only part of this adaptation to the milk environment is shared by *S. thermophilus*. Whereas *L. bulgaricus* has lost most of its amino acid biosynthesis capacity, *S. thermophilus* retains its ability to synthesize all amino acids, except histidine (33). This difference may be explained by the fact that *S. thermophilus* does not generally possess an extracellular protease to exploit the rich source of milk proteins. Perhaps *S. thermophilus* either adapted to the milk environment independently or coevolved with the protease producing *L. bulgaricus* but retained an advantage in conserving its capacity for biosynthesis of amino acids.

Central Carbohydrate Metabolism. As a consequence of the (partial) inactivation of several sugar transport and degradation pathways and inactivation of the LacR repressor (22), carbohydrate metabolism in *L. bulgaricus* appears to prefer the milk sugar lactose. After transport and hydrolysis, this organism selectively metabolizes the glucose moiety, because no enzymes are present to use galactose, which is known to accumulate in the culture medium (6, 34). Whereas glucose, fructose, and mannose can also be metabolized, utilization of lactose results in higher growth rates (35). A similar accumulation of galactose in the culture medium is observed when *S. thermophilus* is grown on lactose, albeit for a different reason. In contrast to *L. bulgaricus*, *S. thermophilus* does possess the necessary genes to direct catabolism of galactose, but these genes are generally repressed in the presence of lactose (36).

Although a putative transporter for ribose and a ribokinase are present, *L. bulgaricus* cannot grow on this sugar as the only carbon source (data not shown), presumably as a consequence of an incomplete pentose-phosphate pathway (see below). A complete glycolytic pathway is present leading to the production of pyruvate, which can subsequently be converted to L- or D-lactate to regenerate NAD⁺.

Three particular features are noted relating to carbohydrate

metabolism (Fig. 12, which is published as supporting information on the PNAS web site).

The first substrate level phosphorylation step of glycolysis, in which glyceraldehyde-3-P is converted to 3-P-glycerate via 1,3-bi-P-glycerate, yielding NADH and ATP, can be bypassed by a nonphosphorylating NADP⁺-dependent glyceraldehyde-3P dehydrogenase (GapN). This would result in glycolysis yielding pyruvate and NADPH with no net ATP production. The GapN protein is not present in other sequenced *Lactobacillus* species but is very common among streptococci. In *Streptococcus mutans*, the role of GapN has been suggested to be the production of NADPH because this bacterium lacks the NADPH-generating enzymes in the oxidative branch of the pentose phosphate pathway (37). GapN may also be implicated in an additional level of regulation of glycolysis, allowing the adjustment of metabolic and energy fluxes (38).

L. bulgaricus contains an apparently intact oxidative, NADPH-generating branch of the pentose phosphate pathway. However, distinct from the related species *L. acidophilus* and *L. johnsonii*, the nonoxidative branch lacks a transketolase. Because pentose-phosphate cycling is generally regarded as a mechanism enabling the efficient synthesis of NADPH (39), the interruption of this cycle would support the hypothesis of an NADPH-producing role for GapN under conditions where the oxidative pentose-phosphate pathway is not able to meet the demand.

Similar to *L. acidophilus* and *L. johnsonii*, no pyruvate dehydrogenase complex is present, nor are other enzymes that could generate acetyl-CoA directly from pyruvate under anaerobic conditions. The most likely alternative route would convert the pentose phosphate pathway intermediate D-xylulose-5-P to acetyl-P, which can subsequently be used to generate acetyl-CoA. Acetyl-P may also be produced from acetate, which has been found to be an essential component of a chemically defined medium for *L. bulgaricus* (35). One may speculate that acetate fulfills this role under conditions where the pentose-phosphate pathway would be unable to supply sufficient substrates for acetyl-CoA generation.

Acetyl-CoA may subsequently be used for the production of saturated fatty acids. It is not clear whether *L. bulgaricus* would be able to produce unsaturated fatty acids, however, because a FabA or FabM homologue is missing (40). Failure to produce acetyl-CoA or the inability to produce unsaturated fatty acids may explain the need for Tween 80 in laboratory media for *L. bulgaricus*. Oleic acid, the unsaturated fatty acid in Tween 80, is also available in milk.

Protocooperation Between *L. bulgaricus* and *S. thermophilus*. *L. bulgaricus* and *S. thermophilus* have long been known to stimulate each other's growth and product acidification during milk fermentation in a process called protocooperation. Several factors have been shown or postulated to be responsible for this effect, of which the most obvious is that *L. bulgaricus* possesses an extracellular cell wall-bound proteinase (32) which, through degradation of milk proteins, can supply peptides and amino acids to *S. thermophilus*, which generally does not possess such a protease. Among the other protocooperative factors are formate and CO₂ produced by *S. thermophilus*, which stimulate the growth of *L. bulgaricus* (42).

Analysis of the *L. bulgaricus* genome revealed additional factors that may play a role in protocooperation. (i) The genome sequence encodes a full set of genes for the biosynthesis of folate (Fig. 13, which is published as supporting information on the PNAS web site), a cofactor in many metabolic reactions and an essential component of the human diet (42). *L. bulgaricus* does not, however, possess a means to produce *p*-aminobenzoic acid (PABA) which feeds into this pathway. *S. thermophilus* does possess the necessary enzymes to produce PABA and folate, and as a consequence yogurt is generally regarded as a source of PABA. *L. bulgaricus* may thus benefit from elevated levels of PABA (and folate) when cocultured with *S. thermophilus*. (ii) Two features in the genome suggest that *L. bulgaricus* has a particular need for polyamines, which participate in many cellular processes and may play a role in tolerance to

oxidative stress (43, 44). First, two ABC transporters are found that could serve the uptake of putrescine and/or spermidine from the culture medium. One of these is encoded by a *potABCD* operon (52% GC) in which the genes are arranged in the same order as found in most other bacteria (ERGO). The other is encoded by atypical, AT-rich (38% GC) genes that are present in the order *potBCAD* found in very few bacteria and probably were acquired by horizontal transfer. Second, two genes are found that encode ornithine decarboxylases that catalyze the conversion of ornithine to putrescine. These enzymes are rare in Gram-positive bacteria, except for lactobacilli of the acidophilus group. Because *L. bulgaricus* does not possess the genes necessary to produce ornithine, one may hypothesize that during yogurt fermentation ornithine is provided by *S. thermophilus*, which does possess the enzymes necessary to produce ornithine, but no ornithine decarboxylase (Fig. 14, which is published as supporting information on the PNAS web site). Interestingly, *S. thermophilus* does also possess an ABC transporter dedicated to the uptake of spermidine/putrescine, and, thus, it could be speculated that *L. bulgaricus* and *S. thermophilus* mutually benefit from an exchange of ornithine and putrescine.

A prediction of protein localization (Tables 11 and 12, which are published as supporting information on the PNAS web site) revealed a limited number of putative cell wall-bound and extracellular proteins. Perhaps some of these could contribute to direct contacts between *L. bulgaricus* and *S. thermophilus* as visualized by Bolotin *et al.* (9). From a fermentation viewpoint, the exopolysaccharide (eps) gene clusters present in *L. bulgaricus* (Table 13, which is published as supporting information on the PNAS web site) and *S. thermophilus* appear highly diverse. Variation of these eps clusters among strains is likely important to the texture characteristics of yogurt and merits further investigation.

Stress Resistance. The *L. bulgaricus* genome sequence reveals few of the genes that are known to be involved in the resistance to oxidative stress or low pH, even though this lactic acid bacterium acidifies the culture medium and can exhibit optimal growth under microaerobic conditions (45). The only enzyme that could eliminate oxygen appears to be pyruvate oxidase, but no catalase is present to detoxify the H₂O₂ produced in this reaction. Strain ATCC11842 does not appear to possess the NADH oxidase described by Marty-Teyssset *et al.* (46) in another *L. bulgaricus* strain. Other genes involved in oxidative stress resistance are two thioredoxins, two thioredoxin reductases, and possibly a homologue of the regulatory RNA polymerase-binding protein Spx (47). The uptake of polyamines or their precursors may also provide an alternative means of improving oxidative stress tolerance (see above). Acid stress caused by the production of lactic acid may be countered primarily by the action of an H⁺ transporting ATPase. The presence of two ornithine decarboxylases and several cation:proton antiporters may also assist in the stabilization of intracellular pH, as in *L. acidophilus* (48).

Conclusions

The *L. bulgaricus* genome sequence reveals a number of features that support the hypothesis of a genome in a rapid phase of evolution. Among the lactobacilli of the acidophilus complex, the species *L. delbrueckii* is often regarded as atypical because of its strongly differing GC content. This difference has been the reason to rename what was originally called the *L. delbrueckii* group of lactobacilli and give it the name of a more representative member, *L. acidophilus* (49). The present genome sequence reveals that the observed difference in GC content is mainly due to a difference in GC content at codon position 3, which may be interpreted as the result of recent evolution and justify the conclusion that *L. bulgaricus* is less atypical than once thought. This conclusion is corroborated by the clear global synteny among the genomes of *L. bulgaricus*, *L. acidophilus*, and *L. johnsonii*, reflecting the presence of a common backbone.

The large number of pseudogenes and other features suggest that the *L. bulgaricus* genome is in a relatively recent state of evolution and adaptation to the dairy environment. For example, it has been reported that *L. bulgaricus* will deplete available folate sources rather than produce folate (50), an essential component of the human diet. The present genome sequence shows that all of the genes of the folate pathway are present and seemingly intact in *L. bulgaricus* ATCC11842.

In the light of a new turn in modern microbiological research with an increasing interest for the metagenome of the human gastrointestinal tract, the functional comparison of the industrial bacterium *L. bulgaricus* and the closely related bacteria from the gastrointestinal tract, *L. acidophilus* and *L. johnsonii*, offers an interesting perspective, particularly because *L. bulgaricus* has evolved away from the other members of the acidophilus complex by its rapid and specialized adaptation to an environment created by man, fermented milk.

Materials and Methods

Sequencing. The genome sequence of the *L. bulgaricus* type strain ATCC11842 was determined by using a shotgun sequencing and assembly strategy followed by multiplex long accurate PCR (51) and gap filling. Two types of libraries with inserts of 5 kbp (obtained by mechanical shearing) and 10 kbp (obtained by enzymatic digestion) were constructed in pcDNA2.1 (Invitrogen) and pBeloBAC11 (California Institute of Technology), respectively. Plasmid insert ends and PCR products were sequenced by using dye-primer and dye-terminator chemistries on LICOR4200L and ABI3700/

ABI3730 sequencers. The genome sequence has been submitted to the European Molecular Biology Laboratory database under accession no. CR954253.

Annotation. A draft annotation of the genome sequence was generated by using self-training gene detection software (SHOW, <http://ssb2.jouy.inra.fr/ssb/SHOW>) and an interface for annotation developed at Institut National de la Recherche Agronomique (Agmial: K.B., V.L., R.B., P.N., S. Chaillou, M.v.d.G., S.P., E.M., M. Hoebeke, P.B., and J.-F.G., unpublished data). This draft annotation was manually checked and was updated where appropriate by using the tools provided in the Institut National de la Recherche Agronomique interface and the ERGO suite developed by Integrated Genomics (52).

Comparative Genome Analysis. Comparison of predicted *L. bulgaricus* proteins with predicted proteins from *L. acidophilus* and *L. johnsonii* and analysis of global synteny between the genomes of these bacteria were performed by using BLAST score ratio analysis software provided by Rasko *et al.* (25).

More details on the methods used are presented as *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site.

We thank A. Sorokin, P. Serror, A. Fernandez, C. Chervaux, and T. Smokvina for useful discussions and N. Galleron, B. Quinquis, V. Brachet, M.-C. Beaussart, J. Musset, and L. Prioux for technical assistance. The work of S.P. and C.G. was in part financed by Danone Vitapole.

- Heller, K. J. (2001) *Am. J. Clin. Nutr.* **73**, 374S–379S.
- Metchnikoff, E. (1907) *The Prolongation of Life* (Heinemann, London).
- Mercenier, A., Pavan, S. & Pot, B. (2003) *Curr. Pharm. Des.* **9**, 175–191.
- Adolfsson, O., Meydani, S. N. & Russell, R. M. (2004) *Am. J. Clin. Nutr.* **80**, 245–256.
- Perdigon, G., Maldonado Galdeano, C., Valdez, J. C. & Medici, M. (2002) *Eur. J. Clin. Nutr.* **56**, Suppl. 4, S21–S26.
- Zourari, A., Accolas, J. P. & Desmazaud, M. J. (1992) *Lait* **72**, 1–34.
- Serror, P., Sasaki, T., Ehrlich, S. D. & Maguin, E. (2002) *Appl. Environ. Microbiol.* **68**, 46–52.
- Ravin, V., Sasaki, T., Räisänen, L., Riipinen, K.-A. & Alatossava, T. (2006) *Plasmid* **55**, 184–193.
- Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S. D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G. D., *et al.* (2004) *Nat. Biotechnol.* **22**, 1554–1558.
- Altermann, E., Russell, W. M., Azcarate-Peril, M. A., Barrangou, R., Buck, B. L., McAuliffe, O., Souther, N., Dobson, A., Duong, T., Callanan, M., *et al.* (2005) *Proc. Natl. Acad. Sci. USA* **102**, 3906–3912.
- Pridmore, R. D., Berger, B., Desiere, F., Vilanova, D., Barretto, C., Pittet, A. C., Zwaalen, M. C., Rouvet, M., Altermann, E., Barrangou, R., *et al.* (2004) *Proc. Natl. Acad. Sci. USA* **101**, 2512–2517.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
- Jansen, R., Embden, J. D., Gastra, W. & Schouls, L. M. (2002) *Mol. Microbiol.* **43**, 1565–1575.
- Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. (2005) *Microbiology* **151**, 2551–2561.
- Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. (2000) *Appl. Environ. Microbiol.* **66**, 1328–1333.
- Condon, C., Liveris, D., Squires, C., Schwartz, I. & Squires, C. L. (1995) *J. Bacteriol.* **177**, 4152–4156.
- Yoshikawa, H. & Ogasawara, N. (1991) *Mol. Microbiol.* **5**, 2589–2597.
- Uchida, T., Ishihara, N., Zenitani, H., Hiratsu, K. & Kinashi, H. (2004) *J. Bacteriol.* **186**, 3313–3320.
- Kuempel, P. L., Henson, J. M., Dircks, L., Tecklenburg, M. & Lim, D. F. (1991) *N. Biol.* **3**, 799–811.
- Aussel, L., Barre, F. X., Aroyo, M., Stasiak, A., Stasiak, A. Z. & Sherratt, D. (2002) *Cell* **108**, 195–205.
- Germond, J. E., Lapiere, L., Delley, M., Mollet, B., Felis, G. E. & Dellaglio, F. (2003) *Mol. Biol. Evol.* **20**, 93–104.
- Lapiere, L., Mollet, B. & Germond, J. E. (2002) *J. Bacteriol.* **184**, 928–935.
- Ravin, V. & Alatossava, T. (2002) *Microbiol. Res.* **157**, 109–114.
- Dupont, L., Boizet-Bonhoure, B., Coddeville, M., Auvray, F. & Ritzenthaler, P. (1995) *J. Bacteriol.* **177**, 586–595.
- Rasko, D. A., Myers, G. S. & Ravel, J. (2005) *BMC Bioinformatics* **6**, 2.
- Brown, M. R. & Kornberg, A. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 16085–16087.
- Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Turchini, R., Peters, S. A., Sandbrink, H. M., Fiers, M. W., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100**, 1990–1995.
- Chastanet, A. & Msadek, T. (2003) *J. Bacteriol.* **185**, 683–687.
- Zuber, U. & Schumann, W. (1994) *J. Bacteriol.* **176**, 1359–1363.
- Teuber, M. (1993) in *Biotechnology*, eds. Rehm, H.-J., Reed, G., Pühler, A. & Stadler, P. (Verlag Chemie, Weinheim, Germany), Vol. 1, pp. 325–366.
- Lerat, E. & Ochman, H. (2004) *Genome Res.* **14**, 2273–2278.
- Gilbert, C., Atlan, D., Blanc, B., Portailier, R., Germond, J. E., Lapiere, L. & Mollet, B. (1996) *J. Bacteriol.* **178**, 3059–3065.
- Hols, P., Hancy, F., Fontaine, L., Grossiord, B., Prozzi, D., Leblond-Bourget, N., Decaris, B., Bolotin, A., Delorme, C., Dusko Ehrlich, S., *et al.* (2005) *FEMS Microbiol. Rev.* **29**, 435–463.
- Welman, A. D. & Maddox, I. S. (2003) *J. Ind. Microbiol. Biotechnol.* **30**, 661–668.
- Chervaux, C., Ehrlich, S. D. & Maguin, E. (2000) *Appl. Environ. Microbiol.* **66**, 5306–5311.
- De Vin, F., Radstrom, P., Herman, L. & De Vuyst, L. (2005) *Appl. Environ. Microbiol.* **71**, 3659–3667.
- Boyd, D. A., Cvitkovitch, D. G. & Hamilton, I. R. (1995) *J. Bacteriol.* **177**, 2622–2627.
- Arutyunov, D. Y. & Muronetz, V. I. (2003) *Biochem. Biophys. Res. Commun.* **300**, 149–154.
- Portais, J. C. & Delort, A. M. (2002) *FEMS Microbiol. Rev.* **26**, 375–402.
- Marrakchi, H., Choi, K. H. & Rock, C. O. (2002) *J. Biol. Chem.* **277**, 44809–44816.
- Drissen, F. M., Kingma, F. & Stadhouders, J. (1982) *Neth. Milk Dairy J.* **36**, 135–144.
- Sybesma, W., Starrenburg, M., Tijsseling, L., Hoefnagel, M. H. & Hugenholtz, J. (2003) *Appl. Environ. Microbiol.* **69**, 4542–4548.
- Igarashi, K. & Kashiwagi, K. (2000) *Biochem. Biophys. Res. Commun.* **271**, 559–564.
- Chattopadhyay, M. K., Tabor, C. W. & Tabor, H. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 2261–2265.
- Schiraldi, C., Adduci, V., Valli, V., Maresca, C., Giuliano, M., Lamberti, M., Carteni, M. & De Rosa, M. (2003) *Biootechnol. Bioeng.* **82**, 213–222.
- Marty-Teyssat, C., de la Torre, F. & Garel, J. (2000) *Appl. Environ. Microbiol.* **66**, 262–267.
- Zuber, P. (2004) *J. Bacteriol.* **186**, 1911–1918.
- Azcarate-Peril, M. A., Altermann, E., Hoover-Fitzula, R. L., Cano, R. J. & Klaenhammer, T. R. (2004) *Appl. Environ. Microbiol.* **70**, 5315–5322.
- Schleifer, K. H. & Ludwig, W. (1995) *Syst. Appl. Microbiol.* **18**, 461–467.
- Crittenden, R. G., Martinez, N. R. & Playne, M. J. (2003) *Int. J. Food Microbiol.* **80**, 217–222.
- Sorokin, A., Lapidus, A., Capuano, V., Galleron, N., Pujic, P. & Ehrlich, S. D. (1996) *Genome Res.* **6**, 448–453.
- Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., *et al.* (2003) *Nucleic Acids Res.* **31**, 164–171.
- Pedersen, A. G., Jensen, L. J., Brunak, S., Staerfeldt, H. H. & Ussery, D. W. (2000) *J. Mol. Biol.* **299**, 907–930.

Supporting information.

Materials and Methods

Sequencing, assembly and quality control. The genome sequence of the *L. bulgaricus* type strain ATCC11842 was determined using a shotgun sequencing and assembly strategy. The Phred/Phrap/Consed software package¹ was used for sequence assembly and quality assessment. The error rate of the assembled sequence was < 1 per 10^5 nucleotides.

Multiplex long accurate PCR² and gap filling were used until one gap remained near the replication terminus. The size of this final gap was estimated by hybridization on chromosomal DNA digests separated by pulsed field gel electrophoresis (PFGE), while PCR analysis learned that this region contained an inverted duplication of the sequence adjacent to the gap. The two copies of the (inverted) repeat sequence were amplified separately to determine their respective sequences as well as the orientation of the unique 1.4 kbp fragment that they enclose. Sequence assemblies across repeated elements (rnn operons and IS elements) were confirmed by PCR and end sequencing of the PCR products. As a further check, chromosomal DNA digests separated by PFGE were hybridized with a 23S rRNA specific probe. The results of this analysis were consistent with the final assembly. Basepair number 1 of the genome map was arbitrarily attributed to the first nucleotide of the *SacII* recognition site situated between the divergent *rpmH* and *dnaA* genes.

Annotation. Short open reading frames (corresponding to proteins < 50 amino acids) were only retained if the coding probability according to the SHOW program was superior to 0.99 and a putative ribosome binding site could be indicated, if blastP or blastX³ results revealed the presence of highly homologous gene products in other bacteria, or if SHOW predicted a highly similar protein in the partially available *Lactobacillus gasseri* genome (<http://jgi.doe.gov/>). The latter criterion was based on the observation that many predicted proteins in *L. bulgaricus* and *L. gasseri* were highly similar while the corresponding DNA sequences were not. tRNA genes were identified using tRNAscan-SE⁴, implemented in the Agmial interface.

Comparative genome analysis. Comparison of predicted *L. bulgaricus* proteins with predicted proteins from *L. acidophilus* and *L. johnsonii* was performed using blast score ratio analysis software provided by Rasko *et al.*⁵. Briefly, for each *L. bulgaricus* protein a blastP raw score was obtained for the alignment against itself (B) and for the alignment against the most similar protein in the second genome (A). The blast score ratio A/B then gives a value between 0 and 1 for each protein. Proteins with a ratio < 0.4 were considered to be nonhomologous. This method overcomes some of the problems that are encountered when blast E-values are used for

comparison, the latter depending on the database used and on the length of the match. The blast score ratio method is less likely to predict a match when only a small region of high similarity (a conserved domain) is shared between two proteins.

¹Ewing, B. & Green, P. (1998) *Genome Res.* **8**, 186-194; Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998) *Genome Res.* **8**, 175-185; Gordon, D., Abajian, C. & Green, P. (1998) *Genome Res.* **8**, 195-202.

²Sorokin, A., Lapidus, A., Capuano, V., Galleron, N., Pujic, P. & Ehrlich, S. D. (1996) *Genome Res.* **6**, 448-453.

³Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403-410.

⁴Lowe, T. M. & Eddy, S. R. (1997). *Nucleic Acids Res.* **25**, 955-964.

⁵Rasko, D. A., Myers, G. S. & Ravel, J. (2005) *BMC Bioinformatics* **6**, 2.

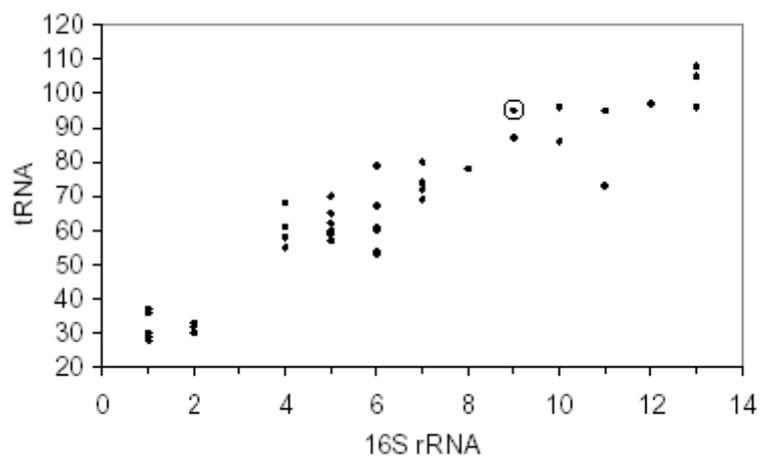


Fig. 5. Relationship between the number of tRNA genes and the number of 16S rRNA genes in 54 firmicutes genomes.

L. bulgaricus (9 16S rRNA, 95 tRNA) data point is circled. For comparison: *L. acidophilus* (4 16S rRNA, 61 tRNA), *L. johnsonii* (6 16S rRNA, 79 tRNA), *L. plantarum* (5 16S rRNA, 70 tRNA).

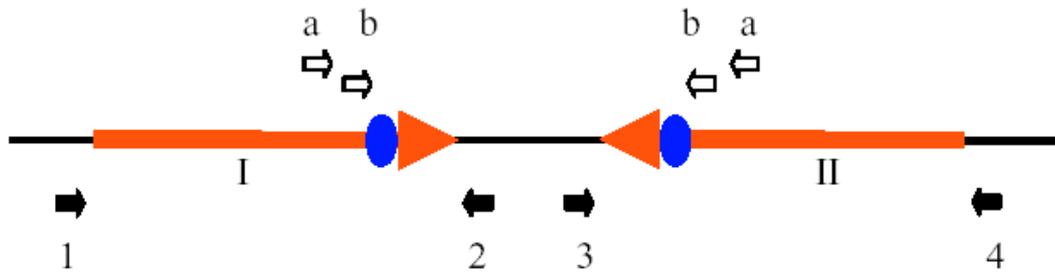


Fig. 6. Inverted repeat (IR) in the replication terminus region of the *L. bulgaricus* genome.

Schematic representation of the IR (not to scale), showing the repeated sequence (red arrows I and II, 23 kbp each) and the duplicated *dif* site (blue). Black lines represent the unique sequence between the repeats (1.4 kbp) and adjacent chromosomal sequences. Arrows a, b, 1, 2, 3, and 4 represent primers used in PCR amplification.

PCR amplifications using only one primer (a or b) indicated the existence of an IR. When the same primers were used on a collection of strains, PCR products of different size were obtained indicating that the size of the unique central sequence and/or the size of the repeat may largely vary between strains. Sequence I could be amplified using primers 1 and 2, or 1 and 3, and sequence II could be amplified using primers 4 and 3, or 4 and 2. Identical repeat sequences were obtained after sequencing of the four PCR products. Although PCR artefacts giving rise to erroneous conclusions can not be excluded, these results suggest that the unique central region between the repeats can be present in either of two orientations, and that sequences I and II form a perfect IR.

B. subtilis : ACTTCCTAGAATATATATTATGTAACT
L. bulgaricus: AGTTCGAATAATACATATTATGTAAAGT

Fig 7. Putative *dif* sequences in *L. bulgaricus* compared to *B. subtilis*¹.

¹ Sciochetti, S. A., Piggot, P. J. & Blakely, G. W. (2001) *J. Bacteriol.* **183**, 1058-1068.

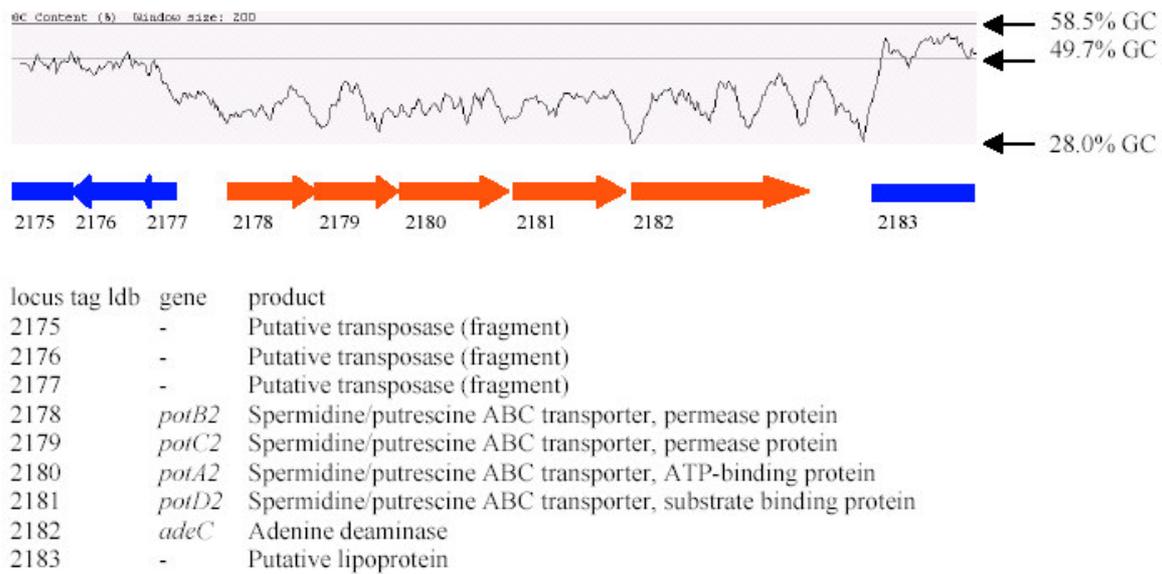


Fig. 8. Example of proximity of a transposase and an AT rich region acquired by horizontal transfer.

Open reading frames are listed by *ldb* gene number. Local GC content (window size 200 bp) is indicated at the top of the figure. Overall genomic GC content is 49.7%. The hypothesis of horizontal transfer of the AT rich *potBCAD-adeC* gene cluster (red arrows) is supported by the fact that this gene order is found in very few, unrelated, bacteria.

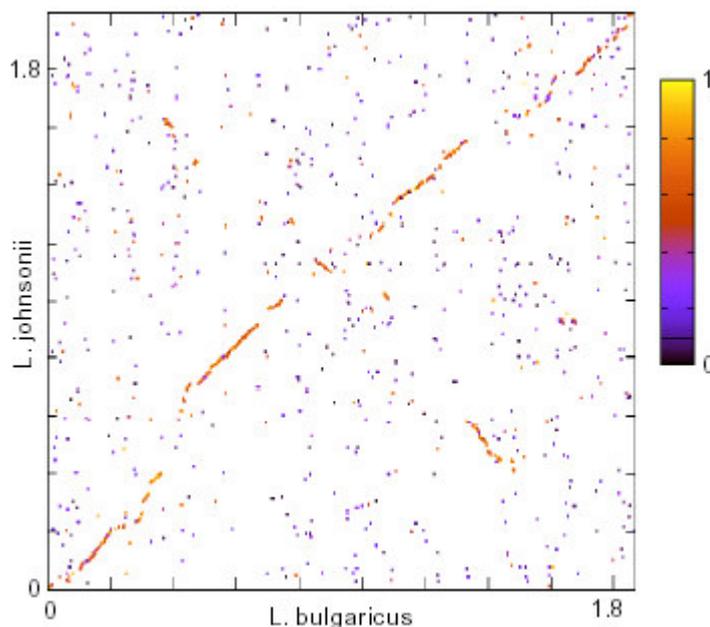


Fig. 9. Synteny between *L. bulgaricus* and *L. johnsonii* genomes.

X-axis, position on *L. bulgaricus* genome (Mbp); Y-axis, position on *L. johnsonii* genome (Mbp). 0, replication origin. Similarity color scale to the right, blast-score ratio¹.

¹ Rasko, D. A., Myers, G. S. & Ravel, J. (2005) *BMC Bioinformatics* **6**, 2.

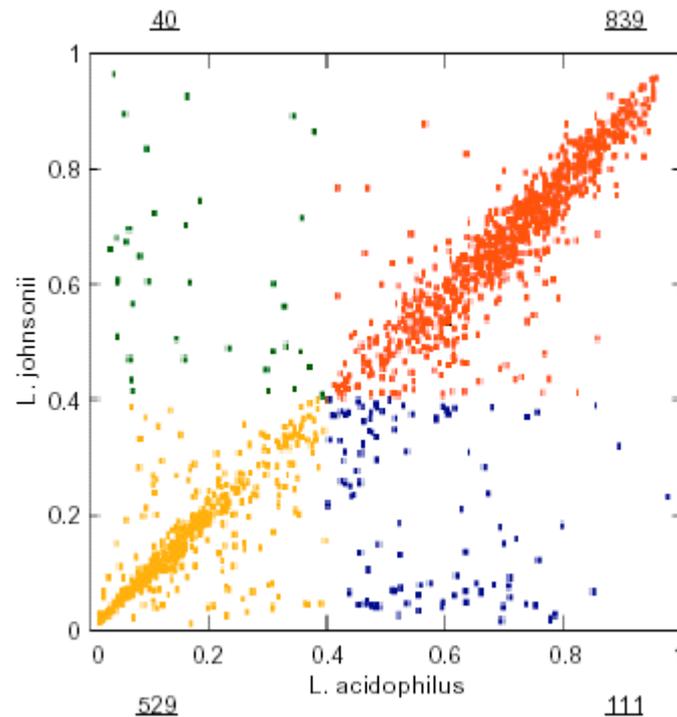


Fig. 10. Resemblance between predicted proteins from *L. bulgaricus* and *L. acidophilus* or *L. johnsonii*.

Predicted proteins from *L. bulgaricus* are positioned according to resemblance (blast score ratio¹) with the closest homologues in *L. johnsonii* and *L. acidophilus*. Red, *L. bulgaricus* proteins that share homology with proteins from *L. acidophilus* and *L. johnsonii*; green, proteins that share homology with proteins from *L. johnsonii* only; blue, proteins that share homology with proteins from *L. acidophilus* only; yellow, proteins that share no homology with proteins from *L. acidophilus* or *L. johnsonii*. Underlined figures indicate the number of proteins in each category. The number of proteins represented in this Figure (1519) is slightly lower than the number of CDS mentioned in Table 1 (1562) because several CDS (located in IS elements or in the large IR described in the text) are predicted to code for identical proteins. See also Table 3.

¹Rasko, D. A., Myers, G. S. & Ravel, J. (2005) *BMC Bioinformatics* **6**, 2.

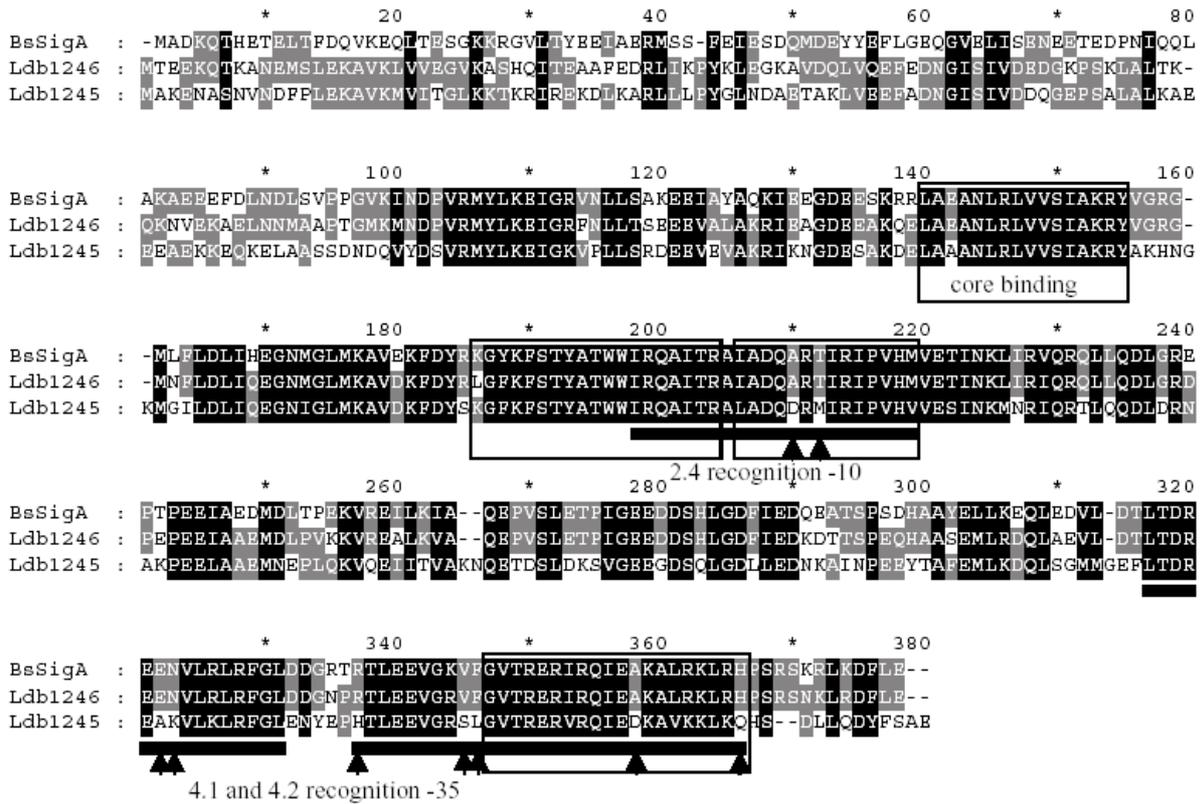


Fig. 11. Alignment of SigA homologues. BsSigA, sigma A of *Bacillus subtilis*.

Ldb1246 and Ldb1245, sigma A homologues of *L. bulgaricus*. Boxed regions are conserved in the sigma 70 family group 1 (sigma A) and group 2 (including *Escherichia coli* sigma S), as opposed to group 3 (alternative sigma factors)¹. Underlined: -35 and -10 promoter recognition domains. Arrowheads indicate non conservative differences between the *L. bulgaricus* homologues in these domains.

¹Lonetto, M., Gribskov, M. & Gross, C. A. (1992) *J. Bacteriol.* **174**, 3843-3849.

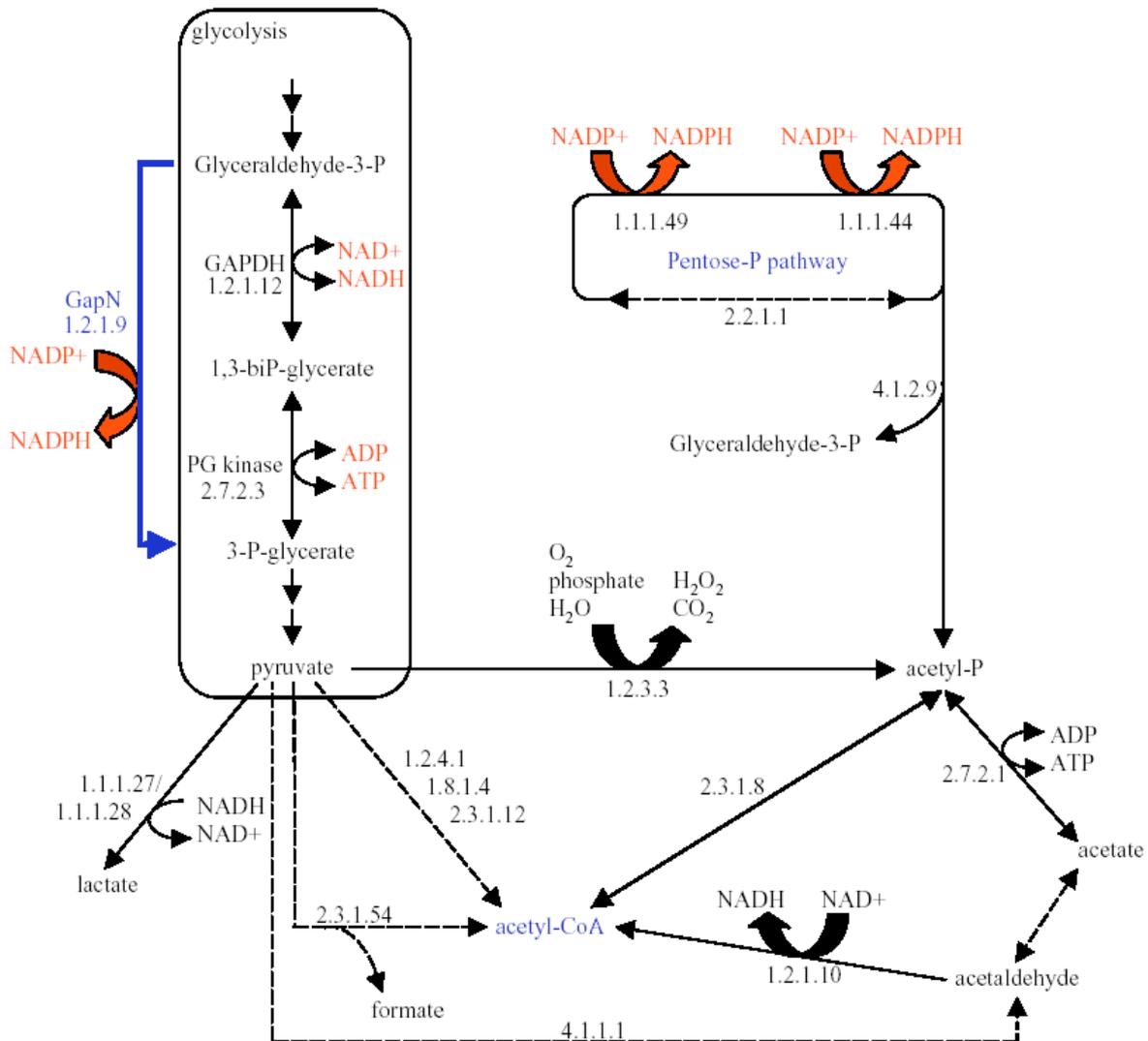


Fig. 12. Particular features of *L. bulgaricus* carbohydrate metabolism.

Illustrated are: GapN bypass in glycolysis, incomplete pentose-P-pathway, pathways for acetyl-CoA formation. Dashed lines indicate the absence of the corresponding genes in the *L. bulgaricus* genome.

1.2.1.12, glyceraldehyde-3-P dehydrogenase (Ldb0635); 2.7.2.3, phosphoglycerate kinase (Ldb0636); 1.1.1.27, L-lactate dehydrogenase (Ldb0120); 1.1.1.28, D- lactate dehydrogenase (Ldb0101); 1.2.1.9, NADP-dependent glyceraldehyde-3-P dehydrogenase (Ldb1179); 1.1.1.49, glucose-6-P 1-dehydrogenase (Ldb1611); 1.1.1.44, phosphogluconate dehydrogenase (Ldb2212); 2.2.1.1, transketolase*; 4.1.2.9, phosphoketolase (Ldb0534); 1.2.3.3, pyruvate oxidase (Ldb2213); 2.3.1.54, pyruvate-formate lyase*; 1.2.4.1, 1.8.1.4, 2.3.1.12, pyruvate dehydrogenase*; 2.3.1.8, phosphate acetyltransferase (Ldb0643); 2.7.2.1, acetate kinase (Ldb0686); 1.2.1.10, acetaldehyde dehydrogenase (Ldb1707); 4.1.1.1, pyruvate decarboxylase*. In parentheses, corresponding protein in *L. bulgaricus*. *Not present in *L. bulgaricus*

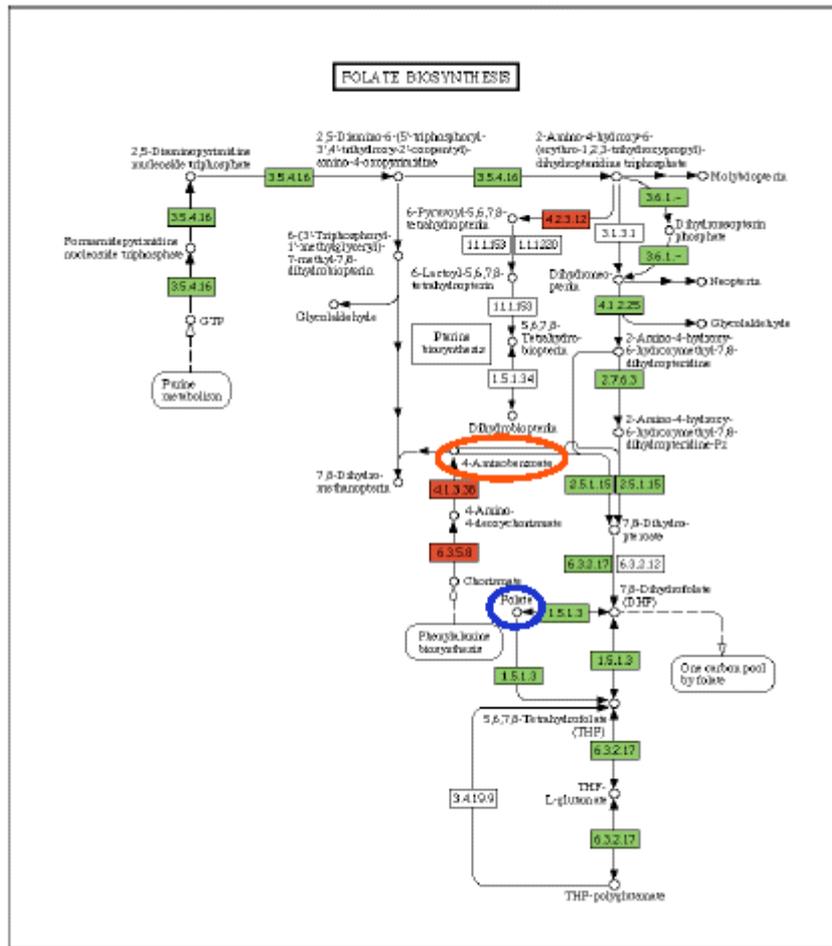


Fig. 13. Proto cooperation and biosynthesis of folate.

Enzymes present in *S. thermophilus* only are indicated in red, enzymes present in both *S. thermophilus* and *L. bulgaricus* ATCC11842 are indicated in green. *S. thermophilus* produces para-aminobenzoic acid (paba, red encircled) which feeds into the folate biosynthesis pathway and may be used by both *S. thermophilus* and *L. bulgaricus* to produce folate (blue encircled). Poly glutamyl derivatives of folate are used as cofactors in the biosynthesis of a variety of cellular compounds¹. Pathway diagram from KEGG², with modifications.

¹McGuire, J.J. & Bertino, J.R. (1981) *Mol. Cell. Biochem.* **38**, 19-48.

²Kanehisa, M. & Goto, S. (2000) *Nucleic Acids Res.* **28**, 27-30.

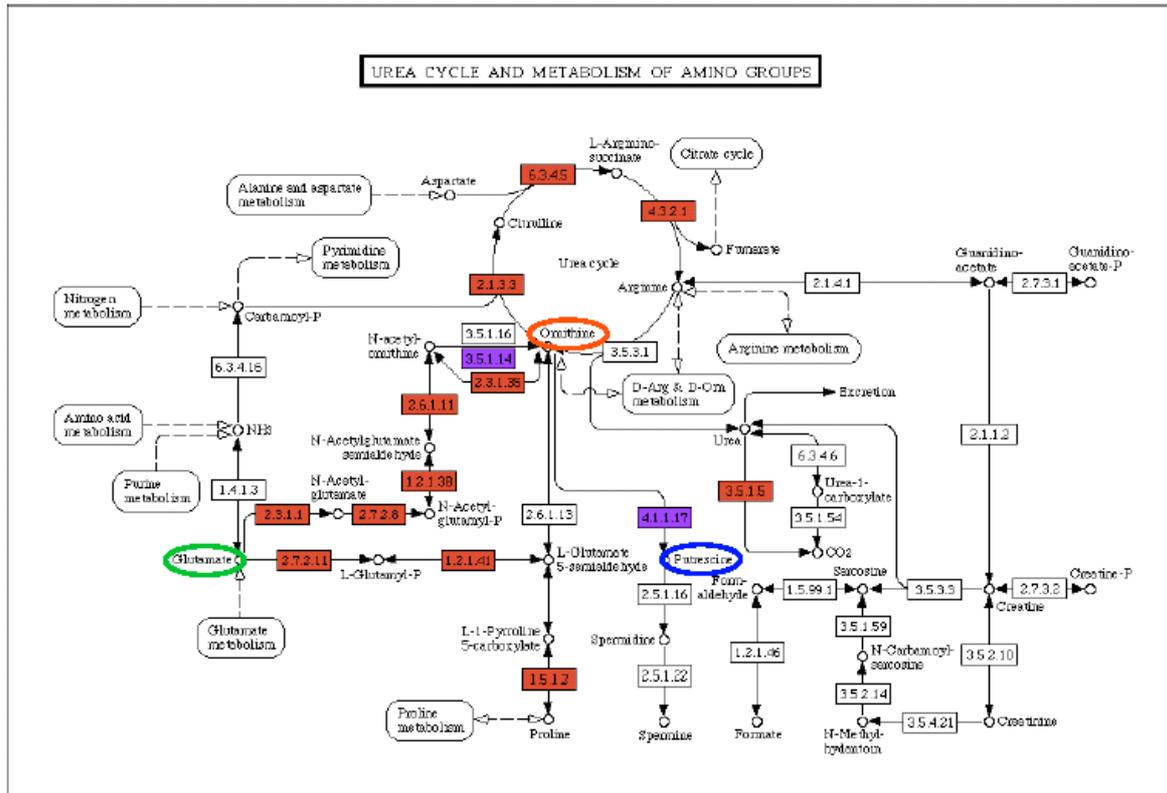


Fig. 14. Protocooperation and biosynthesis of polyamines.

Enzymes present in *S. thermophilus* only are indicated in red, enzymes present in *L. bulgaricus* only are indicated in violet. *S. thermophilus* can convert glutamate (green encircled) to ornithine (red encircled). *L. bulgaricus* can convert ornithine to putrescine (blue encircled). Putrescine may be used by both *L. bulgaricus* and *S. thermophilus* which possesses an ABC transporter dedicated to the uptake of spermidine/putrescine. Pathway diagram from KEGG¹.

¹Kanehisa, M. & Goto, S. (2000) *Nucleic Acids Res.* **28**, 27-30.

Table 2. Transposases in the *L. bulgaricus* genome.

Type	number	of which intact
ISL3	1	0
ISL4	8	8
ISL5	6	0
ISL4-5	6	4
ISL6	1	0
ISL7	16	4
ISLd11	1	0
other	17	3 ^a
total	56	19

Number, total of intact transposase encoding genes and transposase pseudogenes. One additional, partial, ISL4 element was identified on the basis of ISL-4 related hypothetical genes of unknown function.
^abelonging to 2 different types.

Table 3. Proteins encoded in *L. bulgaricus*, not present in *L. acidophilus* or *L. johnsonii*.**A. 374 unique proteins by blast score ratio (bsr < 0.4) and blastP reciprocal best hit.**

function ^a	locus tag	ldb	gene	ec-number	product
1.1	2006		<i>epsIF</i>	-	EpsIF, Hypothetical protein
1.1	2004		<i>epsIH</i>	-	EpsIH, Putative glycosyltransferase
1.1	1951		<i>epsIIC</i>	-	EpsIIC, Putative glycosyltransferase
1.1	1943		<i>epsII-I</i>	-	EpsII-I, Putative glycosyltransferase
1.1	1942		<i>epsIIJ</i>	-	EpsIIJ, Hypothetical membrane protein
1.1	1938		<i>epsIIM</i>	-	EpsIIM, Hypothetical protein
1.1	1998		<i>epsIM</i>	-	EpsIM, Putative glycosyltransferase
1.1	1931		<i>gtcA</i>	-	Teichoic acid glycosylation protein
1.2	1939		<i>epsIIL</i>	-	EpsIIL, Putative polysaccharide repeat unit transport protein
1.2	2178		<i>potB2</i>	-	Spermidine/putrescine ABC transporter, permease protein
1.2	2179		<i>potC2</i>	-	Spermidine/putrescine ABC transporter, permease protein
1.2	2181		<i>potD2</i>	-	Spermidine/putrescine ABC transporter, substrate binding protein
1.2	21		-	-	ABC transporter, ATP-binding protein
1.2	331		-	-	ABC transporter, ATP-binding protein
1.2	460		-	-	ABC transporter, ATP-binding protein
1.2	973		-	-	ABC transporter, ATP-binding protein
1.2	974		-	-	ABC transporter, ATP-binding protein
1.2	1714		-	-	ABC transporter, ATP-binding protein
1.2	1828		-	-	ABC transporter, ATP-binding protein
1.2	2208		-	-	ABC transporter, ATP-binding protein
1.2	14		-	-	ABC transporter, ATP-binding/permease protein
1.2	1288		-	-	ABC transporter, ATP-binding/permease protein
1.2	155		-	-	ABC transporter, permease protein
1.2	156		-	-	ABC transporter, permease protein
1.2	330		-	-	ABC transporter, permease protein
1.2	975		-	-	ABC transporter, permease protein
1.2	1074		-	-	Putative multidrug efflux protein
1.2	1787		-	-	Putative oxalate:formate antiporter
1.2	104		-	-	Substrate binding protein
1.2.3	1239		-	-	Cation transporting P-type ATPase (probable copper transporter)
1.2.5	1712		-	-	Amino acid permease
1.2.5	2198		-	-	Glutamine ABC transporter, substrate binding / permease protein
1.3	963		<i>phoR</i>	-	Phosphate regulon sensor histidine kinase
1.4	1095		-	1.3.1.6	Putative fumarate reductase (flavoprotein)
1.4	102		-	-	Putative oxidoreductase
1.10	682		<i>comGD</i>	-	Putative Competence protein ComGD
2.1.1	1959		<i>gpsA2</i>	1.1.1.94	Glycerol-3-phosphate dehydrogenase
2.1.1	1158		<i>pgp</i>	3.1.3.18	Putative phosphoglycolate phosphatase
2.1.1	475		-	4.2.1.1	Carbonate dehydratase
2.1.1	1147		-	-	D-isomer specific 2-hydroxyacid dehydrogenase
2.1.1	2021		-	-	D-isomer specific 2-hydroxyacid dehydrogenase
2.1.1	1010		-	-	D-isomer specific 2-hydroxyacid dehydrogenase
2.1.1	76		-	3.1.3.46	Putative fructose-2,6-bisphosphatase
2.1.2	1772		<i>pfk2</i>	2.7.1.11	6-phosphofructokinase
2.2	1750		<i>aspA</i>	4.3.1.1	Aspartate ammonia-lyase
2.2	1351		<i>pheB</i>	5.4.99.5	Chorismate mutase
2.2	807		-	2.6.1.-	Aminotransferase
2.3	2182		<i>adeC</i>	3.5.4.2	Adenine deaminase
2.3	1252		<i>cdd</i>	3.5.4.5	Cytidine deaminase
2.3	1363		-	3.1.3.5	Putative secreted 5'-nucleotidase
2.4	909		<i>accA</i>	6.4.1.2	Acetyl-CoA carboxylase, alpha subunit
2.4	905		<i>accB</i>	6.4.1.2	Acetyl-CoA carboxylase, biotin carboxyl carrier protein
2.4	907		<i>accC</i>	6.3.4.14	Acetyl-CoA carboxylase, biotin carboxylase subunit

2.4	908	<i>accD</i>	6.4.1.2	Acetyl-CoA carboxylase, beta subunit
2.4	902	<i>fabD</i>	2.3.1.39	Malonyl-CoA-[acyl-carrier-protein] transacylase
2.4	904	<i>fabF</i>	2.3.1.41	3-oxoacyl-[acyl-carrier protein] synthase II
2.4	900	<i>fabH</i>	2.3.1.41	3-oxoacyl-[acyl-carrier-protein] synthase III
2.4	910	<i>fabI</i>	1.3.1.9	Enoyl-[acyl-carrier protein] reductase (NADH)
2.4	906	<i>fabZ1</i>	4.2.1.-	Hydroxymyristoyl-[acyl carrier protein] dehydratase
2.4	1733	<i>fabZ2</i>	4.2.1.-	Hydroxymyristoyl-[acyl carrier protein] dehydratase
2.5	911	<i>birA</i>	6.3.4.15	Biotin-[acetyl-CoA-carboxylase] ligase
2.5	246	<i>folB</i>	4.1.2.25	Dihydroneopterin aldolase
2.5	248	<i>folC2</i>	6.3.2.17	Folylpolyglutamate synthase
2.5	247	<i>folKE</i>	2.7.6.3	/ 2-Amino-4-hydroxy-6-hydroxymethyldihydropteridine
			3.5.4.16	pyrophosphokinase / GTP cyclohydrolase I
2.6	523	<i>ppk</i>	2.7.4.1	Polyphosphate kinase
2.6	522	<i>ppx</i>	3.6.1.11	Exopolyphosphatase
3.2	1052	<i>hsdM</i>	3.1.21.3	Type I restriction-modification system, modification subunit
3.2	1051	<i>hsdR</i>	3.1.21.3	Type I restriction-modification system, restriction subunit
3.2	1053	<i>hsdS1</i>	3.1.21.3	Type I restriction-modification system, specificity subunit
3.2	1055	<i>hsdS2</i>	3.1.21.3	Type I restriction-modification system, specificity subunit
3.2	1230	<i>mod</i>	2.1.1.72	Type III restriction-modification system methyltransferase
3.2	474	-	-	Putative restriction endonuclease
3.3	1162	-	-	Putative exonuclease
3.3	1079	-	-	Putative integrase/recombinase
3.3	293	-	-	UmuC-like DNA-repair protein
3.5.1	1880	-	-	Putative anti-sigma factor
3.5.1	1881	-	-	Putative RNA polymerase sigma factor (ECF subfamily)
3.5.2	810	<i>argR</i>	-	arginine repressor
3.5.2	48	-	-	Putative transcriptional regulator
3.5.2	1075	-	-	Putative transcriptional regulator of multidrug efflux pump gene (MerR family)
3.5.2	493	-	-	Transcriptional regulator (AsnC family)
3.5.2	1715	-	-	Transcriptional regulator (GntR family)
3.5.2	115	-	-	Transcriptional regulator (LysR family)
3.5.2	124	-	-	Transcriptional regulator (LysR family)
3.5.2	856	-	-	Transcriptional regulator (LysR family)
3.5.2	1290	-	-	Transcriptional regulator (LysR family)
3.5.2	1683	-	-	Transcriptional regulator (LysR family)
3.5.2	1984	-	-	Transcriptional regulator (LysR family)
3.5.2	2141	-	-	Transcriptional regulator (TetR family)
3.6	930	-	2.1.1.-	tRNA/rRNA methyltransferase
3.8	1086	-	-	Putative phosphotyrosine protein phosphatase
3.8	32	-	-	Putative Ser/Thr protein kinase
3.10	1795	<i>dpp</i>	3.4.-.-	Putative dipeptidyl-peptidase
3.10	617	<i>dppA</i>	-	D-aminopeptidase
3.10	1011	-	3.4.19.1	Putative acylaminoacyl-peptidase
4.1	2117	<i>cspA</i>	-	Cold shock protein A
4.2	587	<i>upk</i>	2.7.1.66	Undecaprenol kinase
4.5	1987	-	-	Putative transposase
5.2	868	<i>cas3</i>	-	CRISPR-associated helicase Cas3
5.2	983	-	-	Hydrolase (HAD superfamily)
5.2	2121	-	-	Hydrolase (HAD superfamily)
5.2	1101	-	-	Putative acetyltransferase
5.2	1157	-	-	Putative acetyltransferase
5.2	1765	-	-	Putative acetyltransferase
5.2	1232	-	-	Putative helicase
5.2	1791	-	-	Putative helicase
5.2	325	-	-	Putative hydrolase (NUDIX family)
5.2	1691	-	-	Putative hydrolase (NUDIX family)

not listed: 266 proteins of unknown function

B. 155 unique proteins by blast score ratio (bsr < 0.4) only.

function ^a	locus tag	ldb	gene	ec-number	product
1.1	2005		<i>epsIG</i>	-	EpsIG, Putative glycosyltransferase
1.1	2003		<i>epsI-I</i>	-	EpsI-I, Hypothetical membrane protein
1.1	1944		<i>epsIIH</i>	-	EpsIIH, Putative glycosyltransferase
1.1	510		<i>murl</i>	5.1.1.3	Glutamate racemase
1.1	1695		<i>rfaA</i>	2.7.7.24	Putative glucose-1-phosphate thymidyltransferase
1.1	1755 ⁺		-	-	Aggregation promoting factor
1.1	1664		-	-	Glycosyltransferase
1.1	225		-	-	Putative amidase
1.2	318 ⁺		<i>mscL</i>	-	Large-conductance mechanosensitive channel
1.2	2180 ⁺		<i>potA2</i>	-	Spermidine/putrescine ABC transporter, ATP-binding protein
1.2	1289 ⁺		-	-	ABC transporter, ATP-binding / permease protein
1.2	154		-	-	ABC transporter, ATP-binding protein
1.2	31 ⁺		-	-	ABC transporter, ATP-binding/permease protein
1.2	111 ⁺		-	-	ABC transporter, ATP-binding/permease protein
1.2	302 ⁺		-	-	ABC transporter, ATP-binding/permease protein
1.2	2094 ⁺		-	-	ABC transporter, ATP-binding/permease protein
1.2	2142		-	-	ABC transporter, ATP-binding/permease protein
1.2	1968		-	-	ABC transporter, substrate-binding protein
1.2	1699		-	-	Permease
1.2	1891 ⁺		-	-	Permease
1.2	1686		-	-	Putative ABC transporter, permease protein
1.2	1448		-	-	Putative permease
1.2	1456 ⁺		-	-	Putative permease
1.2.3	957 ⁺		<i>pstC</i>	-	Phosphate ABC transporter, permease protein
1.2.3	956 ⁺		<i>pstS</i>	-	Phosphate ABC transporter, phosphate binding protein
1.2.4	67 ⁺		<i>glpF</i>	-	Glycerol uptake facilitator
1.2.4	199 ⁺		-	2.7.1.69	PTS system, enzyme IIA component
1.2.5	483 ⁺		<i>brnQ</i>	-	Branched-chain amino acid transport system carrier protein
1.3	981		-	-	Putative signal transduction protein
2.1.1	119		<i>glcK</i>	2.7.1.2	Glucokinase
2.1.1	1201*		<i>lacZ</i>	3.2.1.23	Beta-galactosidase
2.1.1	1036 ⁺		-	3.1.3.46	Putative fructose-2,6-bisphosphatase
2.1.2	1179**		<i>gapN</i>	1.2.1.9	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
2.2	1348 ⁺		<i>asd</i>	1.2.1.11	Aspartate semialdehyde dehydrogenase
2.2	1349 ⁺		<i>ask</i>	2.7.2.4	Aspartate kinase
2.2	1325 ⁺		<i>cysK2</i>	2.5.1.47	Cysteine synthase
2.2	1350 ⁺		-	1.1.1.3	Homoserine dehydrogenase
2.4	903 ⁺		<i>fabG</i>	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
2.4	1922		<i>glpQ</i>	3.1.4.46	Glycerophosphoryl diester phosphodiesterase
2.4	1913 ⁺		-	-	Putative lipase
2.5	249		<i>folP</i>	2.5.1.15	Dihydropteroate synthase
2.5	1818 ⁺		<i>iscU</i>	-	NifU homolog involved in Fe-S cluster formation
2.5	319 ⁺		<i>pncB</i>	2.4.2.11	Nicotinate phosphoribosyltransferase
2.5	1820 ⁺		<i>sufD</i>	-	Conserved protein involved in Fe/S cluster assembly
2.6	960		<i>phoU</i>	-	Phosphate uptake regulator
3.1	837 ⁺		<i>dnaE</i>	2.7.7.7	DNA polymerase III, alpha subunit (DnaE)
3.3	1054		-	-	Integrase/recombinase
3.5.1	66		-	-	Putative RNA polymerase sigma factor
3.5.1	1677		-	-	Putative RNA polymerase sigma factor
3.5.2	159		-	-	Putative transcriptional regulator
3.5.2	308		-	-	Putative transcriptional regulator (Crp family)
3.5.2	153		-	-	Putative transcriptional regulator (AraC family)
3.5.2	314		-	-	Putative transcriptional regulator (MarR family)
3.5.2	1793		-	-	Putative transcriptional regulator (Xre family)
3.5.2	1958		-	-	Putative transcriptional regulator (Xre family)

3.5.2	1182	-	-	Transcriptional regulator (DeoR family)
3.5.2	808	-	-	Transcriptional regulator (LysR family)
3.5.2	26	-	-	Two-component system, response regulator
3.5.2	2068 ⁺	-	-	Putative transcriptional regulator (DeoR family)
3.8	1165	-	2.3.1.-	Acyltransferase
3.8	265	-	3.5.1.14	Aminoacylase
3.10	1189	<i>prtB</i>	3.4.21.96	Proteinase precursor
5.2	2067	-	-	FMN-binding protein
5.2	1007	-	-	Hydrolase (HAD superfamily)
5.2	1622 ⁺	-	-	Putative acetyltransferase
5.2	1741 ⁺	-	-	Putative acetyltransferase
5.2	1822 ⁺	-	-	Putative acetyltransferase
5.2	485	-	-	Putative hydrolase (NUDIX family)

not listed: 87 proteins of unknown function

C. 44 unique proteins by blastP reciprocal best hit only.

function ^a	locus	tag	ldb	gene	ec-number	product
1.2	1088	-	-	-	-	ABC transporter, ATP-binding protein
1.2	1089	-	-	-	-	ABC transporter, permease protein
1.2	2162	-	-	-	-	Putative lipoprotein
1.2	2170	-	-	-	-	Putative lipoprotein
1.2.1	281	-	-	<i>oppAIII</i>	-	Oligopeptide ABC transporter, substrate binding protein
1.2.1	1381	-	-	<i>oppA2</i>	-	Oligopeptide ABC transporter, substrate binding protein
1.2.4	2065	-	-	-	2.7.1.69	PTS system, glucose/glucoside-specific enzyme IIA component
1.2.5	521	-	-	<i>glnH2</i>	-	Glutamine ABC transporter, glutamine-binding protein
1.2.5	2199	-	-	-	-	Glutamine ABC transporter, ATP-binding protein
1.2.5	1358	-	-	-	-	Glutamine-binding protein
2.2	441	-	-	-	2.6.1.57	Aromatic amino acid aminotransferase
2.3	1070	-	-	<i>carA2a</i>	6.3.5.5	Carbamoyl-phosphate synthase small chain
2.3	1137	-	-	<i>carA2a</i>	6.3.5.5	Carbamoyl-phosphate synthase small chain
2.3	1530	-	-	<i>pyrD2</i>	1.3.3.1	Dihydroorotate dehydrogenase
2.3	515	-	-	<i>udk2</i>	2.7.1.48	Uridine kinase
2.4	901	-	-	<i>acpP2</i>	-	Acyl carrier protein (ACP)
3.4	1646	-	-	<i>hIbB</i>	-	Histone-like DNA-binding protein
3.5.1	1245	-	-	-	-	RNA polymerase sigma factor
3.7.2	1004	-	-	<i>asnS</i>	6.1.1.22	Asparaginyl-tRNA synthetase
3.8	288	-	-	<i>pepW</i>	3.4.22.-	Aminopeptidase W
3.8	2096	-	-	-	3.1.3.48	Putative protein tyrosine/serine phosphatase
3.10	1700	-	-	<i>pepZ</i>	3.4.13.9	X-Pro dipeptidase PepZ
4.1	2102	-	-	<i>clpL</i>	-	ATP-dependent Clp protease, ATP-binding subunit
4.5	326	-	-	-	-	Putative Transposase
4.5	100	-	-	-	-	Transposase
4.5	803	-	-	-	-	Transposase
4.5	835	-	-	-	-	Transposase
4.5	919	-	-	-	-	Transposase
4.5	1206	-	-	-	-	Transposase
4.5	1320	-	-	-	-	Transposase
4.5	1466	-	-	-	-	Transposase
4.5	1752	-	-	-	-	Transposase
4.5	1874	-	-	-	-	Transposase
4.5	1907	-	-	-	-	Transposase
4.5	1934	-	-	-	-	Transposase
4.5	1979	-	-	-	-	Transposase
4.5	2153	-	-	-	-	Transposase

not listed: 7 proteins of unknown function

Results are based on a blast score ratio (bsr) analysis¹ with threshold 0.4 and a blastP reciprocal best hit (rbh) analysis. Results obtained with both approaches are presented in Table A. Additional results obtained with one of the two approaches are presented in Tables B and C. Bsr and rbh results may differ because of differences between protein length and length of homologous region, presence of paralogues, and overall similarity level, as exemplified by the following proteins marked in table B: **L. acidophilus* and *L. johnsonii* do not contain a *lacZ* gene, but do contain the functional homologous *lacL* and *lacM* genes that each correspond to a part of the *lacZ* gene. ***L. acidophilus* and *L. johnsonii* do not contain a *gapN* gene, but encode distantly related dehydrogenases. [†]0.3 < bsr < 0.4; these proteins are not considered unique to *L. bulgaricus* if bsr analysis is performed with threshold 0.3 instead of 0.4.

[‡]Functional categories: 1.1, Cell wall; 1.2, Transport/binding proteins and lipoproteins; 1.2.1, Transport/binding of proteins/peptides; 1.2.3, Transport/binding of inorganic ions; 1.2.4, Transport/binding of carbohydrates; 1.2.5, Transport/binding of amino-acids; 1.3, Sensors (signal transduction); 1.4, Membrane bioenergetics (electron transport chain and ATP synthase); 1.10, Transformation/competence; 2.1.1, Specific carbohydrate metabolic pathway; 2.1.2, Main glycolytic pathways; 2.2, Metabolism of amino acids and related molecules; 2.3, Metabolism of nucleotides and nucleic acids; 2.4, Metabolism of lipids; 2.5, Metabolism of coenzymes and prosthetic groups; 2.6, Metabolism of phosphate; 3.1, DNA replication; 3.2, DNA restriction and modification (and repair); 3.3, DNA recombination, and repair; 3.4, DNA packaging and segregation; 3.5.1, Transcription initiation; 3.5.2, Transcription regulation; 3.6, RNA modification; 3.7.2, Aminoacyl-tRNA synthetases; 3.8, Protein modification; 3.10, Protein degradation; 4.1, Adaptation to atypical conditions; 4.2, Detoxification. 4.5, Transposon and IS; 5.2, Protein of unknown function similar to proteins from other organisms.

¹ Rasko, D.A., Myers, G.S. & Ravel, J. (2005) *BMC Bioinformatics* **6**, 2.

Table 4. Transcriptional regulators in the *L. bulgaricus* genome.

family	<i>L. bulgaricus</i> ^a (1,9 Mbp)	<i>L. plantarum</i> ^b (3,3 Mbp)
AraC	2	8
ArsR	1	9
AsnC	1	ND
Crp	3	3
DeoR	4	9
GntR	4	14
LacI	1	15
LexA	1	1
LysR	8	22
LytR	4	4
MarR	3	21
MerR	2	10
TetR	5	25
Two-component systems	5 ^c	13
Xre	3	ND
Other	6	80
total	53	234

^anumbers excluding pseudogenes

^bhttp://www.cmbi.kun.nl/lactobacillus/supplementary/Ntable_s01B.html

Genome sizes are indicated in parentheses

^ccomplete systems only

ND, not determined

Table 5. Regulation of heat shock response in different Gram-positive bacteria.

heat shock genes	regulated by								
	Bacilli	Streptococci		Staphylococci		Lactobacilli			
			LI ^b			<i>pl</i> ^b	<i>bu</i> ^b	<i>ac</i> ^b	<i>jo</i> ^b
I <i>dnaK</i>	HrcA	<i>HrcA</i>	HrcA	HrcA+CtsR	HrcA	HrcA	HrcA	HrcA	HrcA
<i>groESL</i>	HrcA	HrcA+CtsR	HrcA	HrcA+CtsR	HrcA	HrcA	HrcA	HrcA	HrcA
II general stress	σ^B	no σ^B	no σ^B	σ^B	no σ^B	no σ^B	no σ^B	no σ^B	no σ^B
III <i>clp</i>	CtsR	CtsR ^a	CtsR	CtsR	CtsR	HrcA^c	HrcA ^c	HrcA ^c	HrcA ^c

^ain *S. salivarius*, *clpP* is under dual control of CtsR and HrcA.

^bresults from conserved sequence motif search^{1,2}; LI, *Lactococcus lactis*; pl, *L. plantarum*; bu, *L. bulgaricus*; ac, *L. acidophilus*; jo, *L. johnsonii*.

^cas deduced from the presence of CIRCE boxes upstream of *clpP* and *clpE*.

Table adapted from Chastanet, A. & Msadek, T. (2003) *J. Bacteriol.* **185**, 683-687.

¹Zuber, U. & Schumann, W. (1994) *J. Bacteriol.* **176**, 1359-1363.

²Derre, I., Rapoport, G. & Msadek, T. (1999) *Mol. Microbiol.* **31**, 117-131.

Table 6. Pseudogenes without intact paralogues in the *L. bulgaricus* genome.

function*	locus tag	ldb	gene	ec-number	product
1.1	1975		<i>rmlC</i>	5.1.3.13	dTDP-4-dehydrorhamnose 3,5-epimerase
1.2.4	194		<i>malF</i>	-	Maltose ABC transporter, permease component
1.2.5	1865-1868		<i>arcD</i>	-	Arginine/ornithine antiporter
1.10	580-581		<i>coiA</i>	-	Putative competence protein CoiA
1.10	604-605		<i>comFC</i>	-	Competence protein ComFC
1.10	679-680		<i>comGB</i>	-	Putative competence protein ComGB
1.10	1518		-	-	Putative nicking enzyme
2.1.1	922-926		<i>adhE</i>	1.1.1.1 / 1.2.1.10	Alcohol-acetaldehyde dehydrogenase
2.1.1	1774		<i>amy1</i>	3.2.1.1	Putative alpha-amylase
2.1.1	1784-1786		<i>amy2</i>	3.2.1.1	Alpha-amylase
2.1.1	192-193		<i>dexB</i>	3.2.1.70	Putative glucan 1,6-alpha-glucosidase
2.1.1	1211-1212		<i>lacA</i>	2.3.1.18	Galactoside-O-acetyltransferase
2.1.1	496-500		<i>mtlD</i>	1.1.1.17	Mannitol-1-phosphate 5-dehydrogenase
2.1.1	304-305		<i>pbg1</i>	3.2.1.86 / 3.2.1.85	6-phospho-beta-glucosidase / 6-phospho-beta-galactosidase
2.1.1	1243-1244		<i>pbg2</i>	3.2.1.86 / 3.2.1.85	6-phospho-beta-glucosidase / 6-phospho-beta-galactosidase
2.1.1	1563-1568		<i>ppsA</i>	2.7.9.2	Phosphoenolpyruvate synthase
2.1.1	1769-1770		<i>scrB</i>	3.2.1.26	Sucrose-6-phosphate hydrolase
2.1.1	68		-	2.4.1.7	Sucrose phosphorylase
2.1.1	58		-	3.2.1.21	Putative beta-glucosidase
2.1.1	60		-	2.7.1.4	Fructokinase
2.1.1	1329		-	1.1.1.5	Acetoin dehydrogenase
2.1.1	1826		-	3.2.1.122	Maltose-6-phosphate glucosidase
2.1.1	2035		-	1.13.12.4	Putative lactate oxidase
2.1.1	2097-2099		-	2.7.1.30	Glycerol kinase
2.2	864		<i>argH</i>	4.3.2.1	Argininosuccinate lyase
2.2	1534		<i>aroA</i>	2.5.1.19	3-phosphoshikimate 1-carboxyvinyltransferase
2.2	1535		<i>aroC</i>	4.2.3.5	Chorismate synthase
2.2	1536		<i>aroK</i>	2.7.1.71	Shikimate kinase
2.2	1327-1328		<i>cysE</i>	2.3.1.30	Serine O-acetyltransferase
2.2	1264		<i>lysA</i>	4.1.1.20	Diaminopimelate decarboxylase
2.2	1177-1178		<i>metA1</i>	2.3.1.46	Homoserine O-succinyltransferase
2.2	106		<i>metA2a</i>	2.3.1.46	Homoserine O-succinyltransferase
2.2	1457		<i>metA2a</i>	2.3.1.46	Homoserine O-succinyltransferase
2.2	376-377		<i>metY</i>	2.5.1.49	O-acetylhomoserine (thiol)-lyase
2.2	1032-1033		<i>serA</i>	1.1.1.95	Phosphoglycerate dehydrogenase
2.2	1034-1035		<i>serC</i>	2.6.1.52	Phosphoserine aminotransferase
2.2	830-832		-	2.7.2.11	Glutamate 5-kinase
2.4	1670-1673		<i>dxs</i>	2.2.1.7	1-deoxy-D-xylulose 5-phosphate synthase
2.4	109		-	-	Putative Acyl-CoA dehydrogenase
2.5	2128-2129		<i>ribD</i>	-	Riboflavin biosynthesis protein RibD
2.5	2127		-	2.5.1.9	Riboflavin synthase alpha chain
2.5	42		-	2.7.1.35	Pyridoxal kinase
2.5	1480-1481		-	6.3.3.2	5-formyltetrahydrofolate cyclo-ligase
3.2	1227-1229		-	3.1.21.5	Type III restriction-modification system endonuclease
3.5.2	1197-1200		<i>lacR</i>	-	Lac repressor
3.5.2	961-962		<i>phoB</i>	-	Phosphate regulon response regulator
3.5.2	1768		<i>scrR</i>	-	Sucrose operon repressor (LacI family)
3.5.2	495		-	-	Mannitol operon transcriptional regulator
3.10	618-620		-	-	L-aminopeptidase
4.2	569-572		-	-	Putative multicopper oxidase
4.2	964		-	-	Putative bile salt hydrolase

Only pseudogenes to which a (inactivated) function could be attributed are indicated.

*Functional categories: 1.1, Cell wall; 1.2.4, Transport/binding of carbohydrates; 1.2.5, Transport/binding of amino-acids; 1.10, Transformation/competence; 2.1.1, Specific carbohydrate metabolic pathway; 2.2, Metabolism of amino acids and related molecules; 2.4, Metabolism of lipids; 2.5, Metabolism of coenzymes and prosthetic groups; 3.2, DNA restriction and modification (and repair); 3.5.2, Transcription regulation; 3.10, Protein degradation; 4.2 Detoxification.

Table 7. Proteolytic system of *L. bulgaricus*.

locus tag	ldb	gene	ec-number	product
1189		<i>prtB</i>	3.4.21.96	Proteinase precursor
1339		-	3.4.24.-	Putative metalloprotease
1727		<i>sms</i>	-	Putative ATP-dependent serine protease
1730		<i>pepC</i>	3.4.22.40	Aminopeptidase C
511		<i>pepD1</i>	3.4.-.-	Dipeptidase
2063		<i>pepD2</i>	3.4.-	Dipeptidase
2034		<i>pepF</i>	3.4.24.-	Oligoendopeptidase F
287		<i>pepG</i>	3.4.22.-	Aminopeptidase G
1896		<i>pepIP</i>	3.4.11.5	Proline iminopeptidase
556		<i>pepM, map</i>	3.4.11.18	Methionine aminopeptidase
2080		<i>pepN</i>	3.4.11.2	Aminopeptidase N
218		<i>pepO1</i>	3.4.24.-	Endopeptidase
1429		<i>pepP</i>	3.4.11.9	X-Pro dipeptidase PepP
1909		<i>pepPN</i>	3.4.11.5	Prolyl aminopeptidase
1594		<i>pepQ1</i>	3.4.13.9	X-Pro dipeptidase PepQ
1217		<i>pepT1</i>	3.4.11.14	Aminotripeptidase (peptidase T)
371		<i>pepT2</i>	3.4.11.14	Aminotripeptidase (peptidase T)
1746		<i>pepV</i>	3.4.13.3	X-His dipeptidase
288		<i>pepW</i>	3.4.22.-	Aminopeptidase W
1455		<i>pepX</i>	3.4.14.-	X-Pro dipeptidyl-peptidase
1700		<i>pepZ</i>	3.4.13.9	X-Pro dipeptidase PepZ
1795		<i>dpp</i>	3.4.-.-	Putative dipeptidyl-peptidase
617		<i>dppA</i>	-	D-aminopeptidase
594		-	-	Putative peptidase
595		-	-	Putative peptidase
1623		-	-	Putative peptidase
1011		-	3.4.19.1	Putative acylaminoacyl-peptidase
1621		<i>gcp</i>	3.4.24.57	Putative glycoprotein endopeptidase
1175		<i>pcp</i>	3.4.19.3	Pyroglutamyl-peptidase I
383		<i>clpC</i>	-	ATP-dependent Clp protease, ATP-binding subunit
574		<i>clpE</i>	-	ATP-dependent Clp protease, ATP-binding subunit ClpE
2102		<i>clpL</i>	-	ATP-dependent Clp protease, ATP-binding subunit
624		<i>clpP</i>	3.4.21.92	ATP-dependent Clp protease proteolytic subunit
778		<i>clpX</i>	-	ATP-dependent Clp protease ATP-binding subunit clpX
1269		<i>hslU, clpY</i>	-	ATP-dependent protease HslVU (ClpYQ), ATPase subunit
1270		<i>hslV, clpQ</i>	3.4.25.-	ATP-dependent protease HslVU (ClpYQ), peptidase subunit
160		<i>htpX1</i>	3.4.24.-	Protease
1185		<i>htpX2</i>	3.4.24.-	Protease
1555		<i>dacA</i>	3.4.16.4	D-alanyl-D-alanine carboxypeptidase
140		<i>htrA</i>	3.4.21.-	HtrA-like serine protease

In addition, a putative peptidylprolyl isomerase or protease maturation protein (PrsA, PrtM) is found (Ldb1533), and two signal peptidases (signal peptidase I (Ldb1215) and signal peptidase II (Ldb1021)).

Table 8. Peptide and amino acid transport systems of *L. bulgaricus*.

locus tag	ldb	gene	product
276		<i>oppA3II</i>	Oligopeptide ABC transporter, substrate binding protein
281		<i>oppA1II</i>	Oligopeptide ABC transporter, substrate binding protein
282		<i>oppBII</i>	Oligopeptide ABC transporter, permease protein
283		<i>oppCII</i>	Oligopeptide ABC transporter, permease protein
284		<i>oppDII</i>	Oligopeptide ABC transporter, ATP-binding protein
285		<i>oppFII</i>	Oligopeptide ABC transporter, ATP-binding protein
1381		<i>oppA2</i>	Oligopeptide ABC transporter, substrate binding protein
1382		<i>oppA1</i>	Oligopeptide ABC transporter, substrate binding protein
1383		<i>oppC</i>	Oligopeptide ABC transporter, permease protein
1384		<i>oppB</i>	Oligopeptide ABC transporter, permease protein
1385		<i>oppF</i>	Oligopeptide ABC transporter, ATP-binding protein
1386		<i>oppD</i>	Oligopeptide ABC transporter, ATP-binding protein
206	-		Peptide binding protein
370	-		Peptide binding protein
784	-		Peptide binding protein
1060	-		Peptide binding protein
517		<i>glnQ</i>	Glutamine ABC transporter, ATP-binding protein
518		<i>glnH1</i>	Glutamine ABC transporter, glutamine-binding protein
519		<i>glnM</i>	Glutamine ABC transporter, permease protein
520		<i>glnP</i>	Glutamine ABC transporter, permease protein
521		<i>glnH2</i>	Glutamine ABC transporter, glutamine-binding protein
2198	-		Glutamine ABC transporter, substrate binding / permease protein
2199	-		Glutamine ABC transporter, ATP-binding protein
1180	-		Glycine betaine ABC transporter, substrate binding / permease protein
1181	-		Glycine betaine ABC transporter, ATP-binding protein
168	-		Amino Acid ABC transporter, ATP-binding protein
169	-		Amino acid ABC transporter, substrate binding / permease protein
251	-		Amino Acid ABC transporter, permease protein
252	-		Amino acid ABC transporter, permease protein
253	-		Amino acid ABC transporter, ATP-binding protein
254	-		Amino acid ABC transporter, substrate binding protein
1298	-		Amino acid ABC transporter, ATP-binding protein
1299	-		Amino Acid ABC transporter, substrate binding /permease protein
539	-		Amino acid permease
672	-		Amino-acid permease
1712	-		Amino acid permease
1796	-		Amino acid permease
483		<i>brnQ</i>	Branched-chain amino acid transport system carrier protein
1358	-		Glutamine-binding protein

Table 9. Putative cellwall/membrane associated proteins and extracellular proteins encoded in the *L. bulgaricus* genome.

A. Cellwall / Membrane.

function	locus tag ldb	gene	ec-number	product	Psortb	SignalP		TMH MM	LipoP
						NN	HMM		
1.1	1555	<i>dacA</i>	3.4.16.4	D-alanyl-D-alanine carboxypeptidase	U 3,33	Y	Y	0	
1.1	2145	<i>dltD</i>	-	D-alanyl transfer protein DltD	U 2,5	Y	Y	1	
1.1	741	<i>murG</i>	2.4.1.	UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	U 2,5	N	N	0	
1.1	1483	<i>pbp</i>	-	Penicillin-binding protein	M 9,82	Y	N	1	
1.1	1012	<i>pbp1A</i>	-	Penicillin-binding protein 1A	E 9,55	N	N	1	
1.1	1542	<i>pbp2A</i>	-	Penicillin-binding protein 2A	E 9,55	N	N	1	
1.1	1755	-	-	Aggregation promoting factor	U 5,44	Y	Y	1	
1.1	224	-	-	Putative amidase	E 9,73	Y	Y	0	
1.1	225	-	-	Putative amidase	E 9,98	Y	Y	0	
1.1	464	-	-	Putative muramidase	E 9,55	Y	N	1	
1.1	1644	-	-	Putative penicillin binding protein	M 9,75	Y	Y	0	
1.1	1645	-	-	Putative penicillin binding protein	E 9,73	Y	Y	0	
1.1	697	-	-	Putative penicillin binding protein	E 9,55	N	N	0	
1.2	650	<i>potD1</i>	-	Spermidine/putrescine ABC transporter, substratebinding protein	U 3,33	Y	Y	1	
1.2	2181	<i>potD2</i>	-	Spermidine/putrescine ABC transporter, substrate binding protein	U 3,33	Y	Y	0	SpII
1.2	792	-	-	ABC transporter, substrate binding protein	U 2,5	Y	N	0	
1.2	1968	-	-	ABC transporter, substrate-binding protein	U 3,33	Y	Y	0	SpII
1.2	2162	-	-	Putative lipoprotein	M 9,35	Y	Y	0	SpII
1.2	2170	-	-	Putative lipoprotein	U 3,33	Y	Y	0	SpII
1.2	2183	-	-	Putative lipoprotein	U 3,33	Y	Y	0	SpII
1.2	104	-	-	Substrate binding protein	U 3,33	Y	Y	1	
1.2	675	-	-	Substrate binding protein	U 3,33	Y	Y	0	SpII
1.2.1	1382	<i>oppA1</i>	-	Oligopeptide ABC transporter, substrate binding protein	U 4,05	Y	Y	0	SpII
1.2.1	281	<i>oppAIII</i>	-	Oligopeptide ABC transporter, substrate binding protein	W 9,17	N	N	0	
1.2.1	1381	<i>oppA2</i>	-	Oligopeptide ABC transporter, substrate binding protein	U 4,05	N	N	0	
1.2.1	276	<i>oppA3II</i>	-	Oligopeptide ABC transporter, substrate binding protein	W 9,18	Y	Y	0	SpII
1.2.1	206	-	-	Peptide binding protein	W 9,18	Y	Y	0	SpII
1.2.1	370	-	-	Peptide binding protein	W 8,93	Y	Y	0	SpII
1.2.1	784	-	-	Peptide binding protein	W 9,18	Y	Y	0	SpII
1.2.1	1060	-	-	Peptide binding protein	W 9,18	Y	Y	0	SpII
1.2.3	202	<i>phnD</i>	-	Phosphonate ABC transporter, substrate binding protein	U 3,33	Y	Y	0	SpII
1.2.3	956	<i>pstS</i>	-	Phosphate ABC transporter, phosphate binding protein	U 2,5	Y	N	0	
1.2.5	518	<i>glnH1</i>	-	Glutamine ABC transporter, glutamine-binding protein	U 3,33	Y	Y	0	
1.2.5	521	<i>glnH2</i>	-	Glutamine ABC transporter, glutamine-binding protein	M 9,75	Y	Y	0	SpII
1.2.5	254	-	-	Amino acid ABC transporter, substrate binding protein	U 2,5	Y	Y	1	

1.2.5	1358	-	-	Glutamine-binding protein	U 2,5	Y	Y	0	SpII
2.3	1363	-	3.1.3.5	Putative secreted 5'-nucleotidase	W 9,18	Y	Y	0	
3.9	1533	<i>prsA</i> , <i>priM</i>	5.2.1.8	Putative peptidylprolyl isomerase	U 3,33	Y	Y	0	SpII
3.10	1621	<i>gcp</i>	3.4.24.57	Putative glycoprotein endopeptidase	E 8,1	N	N	0	
3.10	1189	<i>priB</i>	3.4.21.96	Proteinase precursor	W 9,93	Y	Y	0	
4.1	140	<i>htrA</i>	3.4.21.	HtrA-like serine protease	W 9,18	N	Y	1	
5.2	323	-	-	Conserved hypothetical protein	U 3,33	Y	Y	0	SpII
5.2	913	-	-	Conserved hypothetical protein	U 3,33	Y	Y	0	SpII
5.2	600	-	-	Putative phosphohydrolase (HD superfamily)	C 8,87	Y	Y	1	SpII
6	1553	-	-	Hypothetical protein	W 9,18	Y	Y	1	
6	2087	-	-	Hypothetical protein	W 9,97	Y	N	1	
6	597	-	-	Hypothetical protein	W 9,17	N	N	1	
6	1404	-	-	Hypothetical protein	W 10	N	N	1	
6	378	-	-	Hypothetical protein	U 2,5	Y	Y	1	SpII
6	229	-	-	Hypothetical protein	U 3,33	Y	Y	0	SpII
6	968	-	-	Hypothetical protein	U 2,5	Y	Y	0	SpII
6	1030	-	-	Hypothetical protein	U 3,33	Y	Y	0	SpII
6	890	-	-	Hypothetical protein	W 9,17	Y	Y	0	
6	2081	-	-	Hypothetical protein	W 9,18	Y	Y	0	
6	849	-	-	Hypothetical protein (LysM domain)	U 2,5	N	N	1	
6	165	-	-	Hypothetical protein (lysM domain)	E 9,98	Y	Y	0	

B. Extracellular.

function	locus tag	gene	ec-number	product	Psortb	SignalP	TMH	LipoP
	ldb					NN HMM	MM	
	5.1	1451	-	Hypothetical protein	U 3,33	Y Y	0	
	5.2	560	-	Conserved hypothetical protein	U 2,5	Y Y	1	
	5.2	1003	-	Conserved hypothetical protein	U 2,5	Y Y	1	
	5.2	1354	-	Conserved hypothetical protein	U 2,5	Y Y	1	
	5.2	1597	-	Conserved hypothetical protein	U 6,46	Y Y	1	
	5.2	1598	-	Conserved hypothetical protein	U 3,33	Y Y	1	
	5.2	1894	-	Conserved hypothetical protein	U 2,5	Y Y	1	
	5.2	1900	-	Conserved hypothetical protein	U 3,33	Y Y	1	
	5.2	2022	-	Conserved hypothetical protein	E 9,73	Y Y	1	
	5.2	2077	-	Conserved hypothetical protein	U 2,5	Y Y	1	
	5.2	2122	-	Conserved hypothetical protein	U 2,5	Y Y	1	
	5.2	117	-	Conserved hypothetical protein	U 2,5	Y Y	0	
	5.2	177	-	Conserved hypothetical protein	U 3,33	Y Y	0	
	5.2	2013	-	Conserved hypothetical protein	E 9,55	Y Y	0	
	5.2	2015	-	Conserved hypothetical protein	E 9,55	Y Y	0	
	5.2	2020	-	Conserved hypothetical protein	E 9,73	Y Y	0	
	6	15	-	Hypothetical protein	U 3,33	Y Y	1	
	6	93	-	Hypothetical protein	U 3,33	Y Y	1	
	6	97	-	Hypothetical protein	U 3,33	Y Y	1	
	6	137	-	Hypothetical protein	U 2,5	Y Y	1	
	6	142	-	Hypothetical protein	U 2,5	Y Y	1	
	6	221	-	Hypothetical protein	U 2,5	Y Y	1	
	6	349	-	Hypothetical protein	U 3,33	Y Y	1	
	6	514	-	Hypothetical protein	U 2,5	Y Y	1	
	6	914	-	Hypothetical protein	U 2,5	Y Y	1	
	6	1221	-	Hypothetical protein	U 2,5	Y Y	1	
	6	1297	-	Hypothetical protein	U 3,33	Y Y	1	

6	1529	-	-	Hypothetical protein	U	3,33	Y	Y	1
6	1547	-	-	Hypothetical protein	U	3,33	Y	Y	1
6	1767	-	-	Hypothetical protein	U	2,5	Y	Y	1
6	1831	-	-	Hypothetical protein	U	3,33	Y	Y	1
6	1920	-	-	Hypothetical protein	U	2,5	Y	Y	1
6	1957	-	-	Hypothetical protein	U	3,33	Y	Y	1
6	2168	-	-	Hypothetical protein	U	2,5	Y	Y	1
6	2205	-	-	Hypothetical protein	U	3,33	Y	Y	1
6	2209	-	-	Hypothetical protein	U	2,5	Y	Y	1
6	2211	-	-	Hypothetical protein	U	3,33	Y	Y	1
6	38	-	-	Hypothetical protein	U	5,44	Y	Y	0
6	184	-	-	Hypothetical protein	U	3,33	Y	Y	0
6	1267	-	-	Hypothetical protein	U	2,5	Y	Y	0
6	1548	-	-	Hypothetical protein	U	2,5	Y	Y	0
6	1751	-	-	Hypothetical protein	U	3,33	Y	Y	0
6	1925	-	-	Hypothetical protein	U	3,33	Y	Y	0
6	1963	-	-	Hypothetical protein	U	5,44	Y	Y	0
6	1976	-	-	Hypothetical protein	U	3,33	Y	Y	0
6	2171	-	-	Hypothetical protein	U	3,33	Y	Y	0

Proteins predicted from the *L. bulgaricus* genome (excluding pseudogenes and IS associated genes) were analyzed with a) PSORTb v.2.0¹, b) SignalP 3.0², c) TMHMM 2.0³, and d) LipoP 1.0⁴. Proteins for which a cytoplasmic localization by a) was not contradicted by any of the other methods are considered cytoplasmic. Proteins for which two or more transmembrane helices are predicted by c) are considered membrane or membrane-associated proteins. Proteins for which one transmembrane helix is predicted by c) are considered membrane associated unless the two methods in b) (NN, neural network, and HMM, hidden Markov model) agree on the presence of a signal peptide, in which case the protein is classified as extracellular. Proteins for which no transmembrane helix is predicted by c) are considered cellwall/membrane associated if a signal is detected by d), extracellular if the two methods in b) agree on the presence of a signal peptide, or cytoplasmic. Exceptions to these general rules are based on considerations of protein function, or a differing classification in a) coupled to a high confidence score. Only putative cellwall/membrane associated proteins and extracellular proteins are listed.

PSORTb predictions: C, cytoplasmic; E, extracellular; M, membrane; W, cell wall; U, unknown; figures represent the confidence score on a scale from 0 to 10. SignalP predictions: Y, signal peptide predicted; N, no signal peptide predicted. TMHMM predictions: numbers indicate the number of transmembrane helices predicted. LipoP predictions: SpII, lipoprotein signal peptide predicted.

*Functional categories: 1.1, Cell wall; 1.2, Transport/binding proteins and lipoproteins; 1.2.1, Transport/binding of proteins/peptides; 1.2.3, Transport/binding of inorganic ions; 1.2.4, Transport/binding of carbohydrates; 1.2.5, Transport/binding of amino-acids; 2.3, Metabolism of nucleotides and nucleic acids; 3.9, Protein folding; 3.10, Protein degradation; 4.1, Adaptation to atypical conditions; 5.1, Protein of unknown function similar to other proteins from the same organism; 5.2, Protein of unknown function similar to proteins from other organisms; 6, Protein of unknown function, without similarity to other proteins.

¹Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M. & Brinkman, F.S.L. (2005) *Bioinformatics* **21**, 617-623.

²Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. (2004) *J. Mol. Biol.* **340**, 783-795.

³Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L.L. (2001) *J. Mol. Biol.* **305**, 567-580.

⁴Juncker, A.S., Willenbrock, H., von Heijne, G., Nielsen, H., Brunak, S. & Krogh, A. (2003) *Protein Sci.* **12**, 1652-1662.

Table 10. Eps clusters in the *L. bulgaricus* genome.

locus tag	ldb gene	ec-number	product
1937	<i>tagD, epsIIN</i>	2.7.7.39	EpsIIN, Glycerol-3-phosphate cytidyltransferase
1938	<i>epsIIM</i>	-	EpsIIM, Hypothetical protein
1939	<i>epsIIL</i>	-	EpsIIL, Putative polysaccharide repeat unit transport protein
1942	<i>epsIIJ</i>	-	EpsIIJ, Hypothetical membrane protein
1943	<i>epsII-I</i>	-	EpsII-I, Putative glycosyltransferase
1944	<i>epsIIH</i>	-	EpsIIH, Putative glycosyltransferase
1945	<i>epsIIG</i>	-	EpsIIG, Putative glycosyltransferase
1949	<i>glf1, epsIIE</i>	5.4.99.9	EpsIIE, UDP-galactopyranose mutase
1950	<i>epsIID</i>	-	EpsIID, Putative glycosyltransferase
1951	<i>epsIIC</i>	-	EpsIIC, Putative glycosyltransferase
1952	<i>epsIIB</i>	-	EpsIIB, Putative glycosyltransferase
1953	<i>epsIIA</i>	-	EpsIIA, Transcriptional regulator of eps operon (LytR family)
1998	<i>epsIM</i>	-	EpsIM, Putative glycosyltransferase
2001	<i>epsIK</i>	-	EpsIK, Putative polysaccharide repeat unit transport protein
2002	<i>glf2, epsIJ</i>	5.4.99.9	EpsIJ, UDP-galactopyranose mutase
2003	<i>epsI-I</i>	-	EpsI-I, Hypothetical membrane protein
2004	<i>epsIIH</i>	-	EpsIIH, Putative glycosyltransferase
2005	<i>epsIIG</i>	-	EpsIIG, Putative glycosyltransferase
2006	<i>epsIF</i>	-	EpsIF, Hypothetical protein
2007	<i>epsIE</i>	-	EpsIE, Putative glycosyltransferase
2008	<i>epsID</i>	-	EpsID, Hypothetical protein
2009	<i>epsIC</i>	-	EpsIC, Hypothetical protein
2010	<i>epsIB</i>	-	EpsIB, Hypothetical protein
2011	<i>epsIA</i>	-	EpsIA, Transcriptional regulator of eps operon (LytR family)

Annexes

Annexe 1 : Classification fonctionnelle

1 Cell envelope and cellular processes

- 1.1 Cell wall
- 1.2 Transport/binding proteins and lipoproteins
 - 1.2.1 Transport/binding of proteins/peptides*
 - 1.2.2 Transport/binding of nucleic acids*
 - 1.2.3 Transport/binding of inorganic ions*
 - 1.2.4 Transport/binding of carbohydrates*
 - 1.2.5 Transport/binding of amino-acids*
 - 1.2.6 Transport/binding of nucleosides, nucleotides, purines and pyrimidines*
- 1.3 Sensors (signal transduction)
- 1.4 Membrane bioenergetics (electron transport chain and ATP synthase)
- 1.5 Mobility and chemotaxis
- 1.6 Protein secretion
- 1.7 Cell division
- 1.8 Sporulation
- 1.9 Germination
- 1.10 Transformation/competence

2 Intermediary metabolism

- 2.1 Metabolism of carbohydrates and related molecules
 - 2.1.1 Specific carbohydrate metabolic pathway
 - 2.1.2 Main glycolytic pathways
 - 2.1.3 TCA cycle
 - 2.1.4 Substrate-specific entries to carbohydrate metabolic pathway*
- 2.2 Metabolism of amino acids and related molecules
- 2.3 Metabolism of nucleotides and nucleic acids
- 2.4 Metabolism of lipids
- 2.5 Metabolism of coenzymes and prosthetic groups
- 2.6 Metabolism of phosphate
- 2.7 Metabolism of sulfur
- 2.8 Metabolism of nitrogen/nitrate and nitrite*

3 Information pathways

- 3.1 DNA replication
- 3.2 DNA restriction and modification °
- 3.3 DNA recombination and repair°
- 3.4 DNA packaging and segregation
- 3.5 RNA synthesis
 - 3.5.1 Transcription initiation
 - 3.5.2 Transcription regulation
 - 3.5.3 Transcription elongation
 - 3.5.4 Transcription termination
- 3.6 RNA modification
- 3.7 Protein synthesis
 - 3.7.1 Ribosomal proteins
 - 3.7.2 Aminoacyl-tRNA synthetases
 - 3.7.3 Translation initiation

3.7.4 Translation elongation

3.7.5 Translation termination

3.7.6 Nonribosomal protein synthesis*

3.8 Protein modification

3.9 Protein folding

3.10 Protein degradation*

4 Other functions

4.1 Adaptation to atypical conditions

4.2 Detoxification

4.3 Antibiotic production

4.4 Phage-related function

4.5 Transposon and IS

4.6 Miscellaneous

5 Proteins of unknown function that are similar to other proteins

5.1 Protein of unknown function similar to other proteins from the same organism

5.2 Protein of unknown function similar to proteins from other organisms

6 Protein of unknown function, without similarity to other proteins

Classification fonctionnelle d'Agmial basée sur la classification de SubtiList (<http://genolist.pasteur.fr/SubtiList/>). * : classes créées ; ° classes modifiées.

Annexe 2 : Classification des CDS de *L. bulgaricus*

Les 248 gènes de la classe II

n°ldb	gène	Produit	CAI	PM théorique (kDa)	PI théorique	DéTECTÉ en 2DE
Métabolisme de la paroi						
1974	<i>rmlD</i>	dTDP-4-dehydrorhamnose reductase	0,589	35,8	4,87	
1755	<i>apf</i>	Aggregation promoting factor	0,573	30,5	8,93	
558	<i>galU</i>	UTP-glucose-1-phosphate uridylyltransferase	0,56	33,9	6,42	x
731	<i>mreB1</i>	Cell-shape-determining protein MreB	0,54	34,8	5,13	x
2146	<i>dltC</i>	D-alanyl carrier protein	0,528	9,3	3,87	
1949	<i>glf1</i> , <i>epsIIE</i>	EpsIIE, UDP-galactopyranose mutase	0,492	43,2	5,47	x
2002	<i>glf2</i> , <i>epsIJ</i>	EpsIJ, UDP-galactopyranose mutase	0,484	43,1	5,75	
2010	<i>epsIB</i>	EpsIB, Hypothetical protein	0,484	30,3	9,68	
0714	<i>mreB2</i>	Cell-shape-determining protein MreB	0,473	34,9	4,95	x
1522	<i>murC</i>	UDP-N-acetylmuramate-alanine ligase	0,465	48,9	5,12	
Transport						
576	<i>ptsH</i>	Phosphocarrier protein HPr	0,804	9,3	4,8	*
1799		PTS system, mannose-specific enzyme IID component	0,698	33,7	9,38	
2170		Putative lipoprotein	0,694	38,6	9,59	
0202	<i>phnD</i>	Phosphonate ABC transporter, substrate binding protein	0,642	34,2	9,85	
0276	<i>oppA3II</i>	Oligopeptide ABC transporter, substrate binding protein	0,638	60,2	9,71	
0577	<i>ptsI</i>	Phosphoenolpyruvate-protein phosphotransferase (HprK)	0,637	63,1	4,75	x
1800		PTS system, mannose-specific enzyme IIC component	0,635	27,9	4,43	
2040		PTS system, fructose-specific enzyme IIABC component	0,593	69,5	5,91	
2183		Putative lipoprotein	0,586	38,2	9,8	
0067	<i>glpF</i>	Glycerol uptake facilitator	0,572	26,2	8,34	
1801		PTS system, mannose-specific enzyme IIAB component	0,566	36,6	5,2	x
0284	<i>oppDII</i>	Oligopeptide ABC transporter, ATP-binding protein	0,562	38,8	5,74	
0792		ABC transporter, substrate binding protein	0,553	30,1	9,97	
1382	<i>oppA1</i>	Oligopeptide ABC transporter, substrate binding protein	0,553	65,9	9,67	x
0285	<i>oppFII</i>	Oligopeptide ABC transporter, ATP-binding protein	0,522	36,2	6,03	x
1771		PTS system, sucrose-specific enzyme IIA component	0,511	15,8	5,16	
0283	<i>oppCII</i>	Oligopeptide ABC transporter, permease protein	0,505	37,5	8,99	
2198		Glutamine ABC transporter, substrate binding / permease protein	0,503	53,6	9,51	
0206		Peptide binding protein	0,502	60,0	9,69	
2200		Permease	0,499	46,2	9,14	
0282	<i>oppBII</i>	Oligopeptide ABC transporter, permease protein	0,497	33,3	9,01	
0203	<i>phnC</i>	Phosphonate ABC transporter, ATP binding protein	0,494	28,6	9,08	
2162		Putative lipoprotein	0,49	37,9	9,87	
0204	<i>phnB</i>	Phosphonate ABC transporter, permease protein	0,483	29,4	8,75	
2085	<i>citT</i>	2-oxoglutarate/malate translocator	0,483	50,7	9,36	
1787		Putative oxalate:formate antiporter	0,482	43,8	8,81	
1968		ABC transporter, substrate-binding protein	0,475	34,8	9,17	
0195	<i>malG</i>	Maltose ABC transporter, permease component	0,474	29,8	9,21	
0503		ABC transporter, ATP-binding protein	0,471	60,9	4,83	
0205	<i>phnE</i>	Phosphonate ABC transporter, permease protein	0,469	29,0	9,52	
0957	<i>pstC</i>	Phosphate ABC transporter, permease protein	0,463	32,1	9,61	
Transport d'électrons et ATP synthase						
0706	<i>atpE</i>	H ⁺ transporting ATPase / ATP synthase, C subunit	0,69	7,6	8,25	
1602	<i>trxA1</i>	Thioredoxin	0,601	11,7	4,28	

0795		Putative fumarate reductase (flavoprotein)	0,581	49,3	5,52	
2060		Putative oxidoreductase	0,549	19,6	4,79	x
0711	<i>atpD</i>	H ⁺ transporting ATPase / ATP synthase, beta subunit	0,513	52,2	4,74	x
2061		Putative oxidoreductase	0,485	19,4	4,9	
0709	<i>atpA</i>	H ⁺ transporting ATPase / ATP synthase, alpha subunit	0,473	54,6	5,22	x
0710	<i>atpG</i>	H ⁺ transporting ATPase / ATP synthase, gamma subunit	0,465	35,4	6,03	
Métabolisme carboné						
1294	<i>eno</i>	enolase	0,835	46,3	4,88	x
0635	<i>gap</i>	Glyceraldehyde 3-phosphate dehydrogenase	0,822	36,6	5,51	x
0101	<i>ldhA</i>	D-lactate dehydrogenase	0,816	37,0	4,88	x
0230	<i>gpmA1</i>	Phosphoglycerate mutase	0,813	26,1	5,24	x
1544	<i>fba</i>	Fructose-bisphosphate aldolase	0,813	33,1	4,9	x*
0636	<i>pgk</i>	Phosphoglycerate kinase	0,766	42,7	5,28	x
2213	<i>pox1</i>	Pyruvate oxidase	0,763	67,7	4,96	*
0839	<i>pyk</i>	pyruvate kinase	0,76	62,9	5,27	x
0637	<i>tpi</i>	Triosephosphate isomerase	0,744	27,4	4,74	x
2086	<i>citG</i>	Fumarate hydratase	0,728	50,9	4,94	
0692	<i>pgi</i>	Glucose-6-phosphate isomerase	0,671	49,3	4,82	x
1792	<i>galE</i>	UDP-glucose 4-epimerase	0,582	35,9	5,84	x
1773	<i>pgm</i>	Phosphoglucomutase	0,561	63,6	5,01	x
0119	<i>glcK</i>	Glucokinase	0,558	32,5	4,92	x
0686	<i>ack</i>	Acetate kinase	0,538	43,2	5,05	x
0534		phosphoketolase	0,525	90,9	5,15	
1959	<i>gpsA2</i>	Glycerol-3-phosphate dehydrogenase	0,514	36,1	5,36	x
2212	<i>gnd</i>	Phosphogluconate dehydrogenase (decarboxylating)	0,51	51,2	4,91	
2041	<i>fruK</i>	1-phosphofructokinase	0,485	32,5	4,63	x
1010		D-isomer specific 2-hydroxyacid dehydrogenase	0,484	36,8	4,96	
0838	<i>pfk1</i>	6-phosphofructokinase	0,48	34,0	5,11	x
1772	<i>pfk2</i>	6-phosphofructokinase	0,479	38,8	9,22	
1179	<i>gapN</i>	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	0,47	51,0	5,24	x
0654	<i>manB</i>	Phosphomannomutase	0,465	48,9	4,96	
Métabolisme des acides aminés						
1472	<i>glnA</i>	Glutamine synthetase (Glutamate-ammonia ligase)	0,639	50,2	5,07	x
0494	<i>glmS</i>	Glutamine-fructose-6-phosphate transaminase (isomerizing)	0,566	65,7	5,13	x
1305	<i>ilvE</i>	Branched-chain amino acid aminotransferase	0,544	37,9	5,15	
0116	<i>metE</i>	Putative methionine synthase	0,524	42,1	4,72	
Métabolisme des acides nucléiques						
0417	<i>adk</i>	Adenylate kinase	0,593	24,2	5,29	x
1600		Xanthosine triphosphate pyrophosphatase	0,557	22,3	5,42	x
0350	<i>prs1</i>	Ribose-phosphate pyrophosphokinase	0,556	35,5	6,54	x
1442	<i>purS</i>	Phosphoribosylformylglycinamide (FGAM) synthase, PurS component	0,554	9,3	4,79	
0198	<i>ndt1</i>	Nucleoside deoxyribosyltransferase	0,551	18,1	4,65	
1443	<i>purC</i>	Phosphoribosylaminoimidazole-succinocarboxamidesynthetase	0,543	27,7	5	x
1444	<i>purB</i>	Adenylosuccinate lyase	0,54	48,9	5,76	x
0538	<i>hprT</i>	hypoxanthine-guanine phosphoribosyltransferase	0,536	21,1	4,9	
0291	<i>purA</i>	Adenylosuccinate synthetase	0,524	47,0	5,48	
1440	<i>purL</i>	Phosphoribosylformylglycinamide synthase II	0,524	80,0	4,7	x
1436	<i>purH</i>	Phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase	0,523	56,6	5,25	x
0297	<i>guaB</i>	Inosine-5 -monophosphate dehydrogenase	0,52	40,1	5,42	
0354	<i>pyrG</i>	CTP synthase	0,516	60,8	5,51	x
0211	<i>nrdD</i>	Anaerobic ribonucleoside-triphosphate reductase	0,506	83,7	5,8	
1343	<i>pyrH</i>	Uridylate kinase	0,503	25,9	7,83	x

1532	<i>pyrE</i>	Orotate phosphoribosyltransferase	0,502	23,3	5,36	x
0704	<i>upp</i>	Uracil phosphoribosyltransferase	0,501	22,8	5,99	x
1308	<i>apt</i>	Adenine phosphoribosyltransferase	0,473	19,8	5,14	x
0298	<i>guaA</i>	GMP synthase	0,455	57,4	5,26	x
Métabolisme des lipides						
1387	<i>acpP1</i>	Acyl carrier protein (ACP)	0,637	8,9	3,98	
0901	<i>acpP2</i>	Acyl carrier protein (ACP)	0,612	9,2	3,77	
0052		4-oxalocrotonate tautomerase	0,509	7,0	5,22	
Modification et dégradation des protéines						
1730	<i>pepC</i>	Aminopeptidase C	0,662	50,9	4,91	x
1104	<i>ppiB</i>	peptidyl-prolyl cis-trans isomerase (rotamase),cyclophilin type	0,615	21,7	4,91	x
2080	<i>pepN</i>	Aminopeptidase N	0,55	95,4	4,74	
0287	<i>pepG</i>	Aminopeptidase G	0,529	49,6	5,13	x
1217	<i>pepT1</i>	Aminotripeptidase (peptidase T)	0,517	46,3	4,61	x
1746	<i>pepV</i>	X-His dipeptidase	0,514	52,0	4,56	x
1594	<i>pepQ1</i>	X-Pro dipeptidase PepQ	0,472	41,1	4,86	x
Transcription						
0353	<i>rpoE</i>	DNA-directed RNA polymerase, delta subunit	0,593	21,8	3,75	
0387	<i>rpoC</i>	DNA-directed RNA polymerase beta' chain	0,58	136,3	6,06	
1667	<i>nusG</i>	Transcription antitermination protein nusG	0,566	20,9	4,71	x
0386	<i>rpoB</i>	DNA-directed RNA polymerase beta chain	0,56	135,8	4,93	x
0422	<i>rpoA</i>	DNA-directed RNA polymerase, alpha chain	0,555	34,8	4,83	x
1414	<i>rpoZ</i>	DNA-directed RNA polymerase, omega chain	0,549	7,8	4,73	
1485	<i>greA1</i>	Transcription elongation factor GreA	0,495	18,0	4,64	
0688		Two-component system, response regulator	0,491	27,9	5,42	x
1593	<i>ccpA</i>	Catabolite control protein A (CcpA)	0,482	36,9	5,82	x
0856		Transcriptional regulator (LysR family)	0,479	35,9	5,76	x
1335	<i>nusA</i>	Transcription elongation factor NusA	0,473	44,6	4,3	
1493		Two-component system, response regulator	0,473	27,2	6,87	
Traduction						
1657	<i>rplL</i>	50S ribosomal protein L7/L12	0,834	12,3	4,51	x
776	<i>tuf</i>	Elongation factor Tu	0,832	43,3	4,85	x
1430	<i>rpmA</i>	50S ribosomal protein L27	0,818	10,9	10,22	
0773	<i>rpsO</i>	30S ribosomal protein S15	0,812	10,5	9,1	
0722	<i>rpsD</i>	30S ribosomal protein S4	0,804	23,5	9,81	
1431	<i>rplU</i>	50S ribosomal protein L21	0,775	11,3	9,35	
0399	<i>rplB</i>	50S ribosomal protein L2	0,77	30,3	10,46	
0429	<i>rpsI</i>	30S ribosomal protein S9	0,769	14,3	10,35	
0393	<i>rpsG</i>	30S ribosomal protein S7	0,766	18,0	9,73	
1371	<i>rpsP</i>	30S ribosomal protein S16	0,756	10,5	9,93	
1666	<i>rplK</i>	50S ribosomal protein L11	0,755	14,8	9,56	
1665	<i>rplA</i>	50S Ribosomal protein L1	0,749	24,6	9,11	
0394	<i>fusA</i>	Translation elongation factor G (EF-G)	0,748	76,6	4,84	x
0428	<i>rplM</i>	50S ribosomal protein L13	0,747	16,4	9,55	
0408	<i>rplE</i>	50S ribosomal protein L5	0,745	20,1	9,42	
1345	<i>rpsB</i>	30S ribosomal protein S2	0,739	28,4	4,98	x
1500	<i>rplT</i>	50S ribosomal protein L20	0,737	13,5	11,13	
0404	<i>rpmC</i>	50S ribosomal protein L29	0,736	7,7	9,45	
0410	<i>rpsH</i>	30S ribosomal protein S8	0,73	14,5	9,15	
1344	<i>tsf</i>	Translation elongation factor Ts	0,73	37,4	4,99	x
0007	<i>rpsF</i>	30S ribosomal protein S6	0,725	11,2	5,21	x
0405	<i>rpsQ</i>	30S ribosomal protein S17	0,719	10,5	10,09	
0420	<i>rpsM</i>	30S ribosomal protein S13	0,713	13,2	10,4	
0402	<i>rpsC</i>	30S ribosomal protein S3	0,712	24,9	10,27	
0392	<i>rpsL</i>	30S ribosomal protein S12	0,711	14,9	10,94	

0772	<i>rpsT</i>	30S ribosomal protein S20	0,711	9,0	10,71	
1658	<i>rplJ</i>	50S ribosomal protein L10	0,706	18,2	4,99	x
0851	<i>rpsA</i>	30S ribosomal protein S1	0,705	44,2	5,42	x
0421	<i>rpsK</i>	30S ribosomal protein S11	0,704	13,7	11,27	
0413	<i>rpsE</i>	30S ribosomal protein S5	0,703	18,0	10,04	
0415	<i>rplO</i>	50S ribosomal protein L15	0,702	15,6	10,39	
2218	<i>rpmH</i>	50S ribosomal protein L34	0,7	5,4	12,61	
0423	<i>rplQ</i>	50S ribosomal protein L17	0,696	14,4	10,1	
1501	<i>rpmI</i>	50S Ribosomal protein L35	0,694	7,8	12,66	
1364	<i>rplS</i>	50S ribosomal protein L19	0,693	13,2	11,2	
0356	<i>rpmE</i>	50S ribosomal protein L31	0,692	9,4	9,07	
0406	<i>rplN</i>	50S ribosomal protein L14	0,679	13,1	10,12	
0403	<i>rplP</i>	50S ribosomal protein L16	0,673	16,3	10,41	
0398	<i>rplW</i>	50S ribosomal protein L23	0,671	11,3	9,43	
1506	<i>thrS</i>	Threonyl-tRNA synthetase	0,653	72,9	5,05	
0411	<i>rplF</i>	50S ribosomal protein L6	0,649	19,3	9,66	
0397	<i>rplD</i>	50S ribosomal protein L4	0,648	22,6	10,2	
0400	<i>rpsS</i>	30S ribosomal protein S19	0,642	10,3	10,06	
0822	<i>rpmF</i>	50S ribosomal protein L32	0,636	7,2	10,34	
0395	<i>rpsJ</i>	30S ribosomal protein S10	0,624	11,6	9,34	
0396	<i>rplC</i>	50S ribosomal protein L3	0,623	22,8	9,87	
0009	<i>rpsR</i>	30S ribosomal protein S18	0,611	8,9	10,75	
0606	<i>raiA</i>	Ribosome associated inhibitor	0,61	21,9	5,34	*
0407	<i>rplX</i>	50S ribosomal protein L24	0,603	8,2	9,8	
1685	<i>gltX</i>	Glutamyl-tRNA synthetase	0,594	57,1	5,38	
1669	<i>rpmG2</i>	50S ribosomal protein L33	0,593	5,5	10,14	
1882	<i>serS</i>	Seryl-tRNA synthetase	0,58	49,3	5,11	x
0809	<i>argS</i>	Arginyl-tRNA synthetase	0,576	64,2	5,23	x
1607	<i>alaS</i>	Alanyl-tRNA synthetase	0,572	96,1	4,97	
0470	<i>gatB</i>	Glutamyl-tRNA(Gln) amidotransferase subunit B	0,571	53,6	5,92	
0373	<i>lysS</i>	Lysyl-tRNA synthetase	0,561	57,7	5,2	x
1570	<i>leuS</i>	Leucyl-tRNA synthetase	0,555	92,0	5,37	x
1428	<i>efp1</i>	Elongation factor P (EF-P)	0,544	21,0	4,88	
0469	<i>gatA</i>	Glutamyl-tRNA(Gln) amidotransferase subunit A	0,543	51,9	4,82	x
1248	<i>glyS</i>	Glycyl-tRNA synthetase, beta subunit	0,54	78,7	5,13	x
0011	<i>rplI</i>	50S ribosomal protein L9	0,533	16,9	9,6	
0342	<i>metS</i>	Methionyl-tRNA synthetase	0,531	75,2	6,11	x
0748	<i>ileS</i>	Isoleucyl-tRNA synthetase	0,529	106,0	5,17	
1911	<i>efp2</i>	Elongation factor P (EF-P)	0,52	20,3	4,85	
0468	<i>gatC</i>	Glutamyl-tRNA(Gln) amidotransferase subunit C	0,515	11,6	4,42	
0727	<i>valS</i>	Valyl-tRNA synthetase	0,515	100,9	5,25	x
1004	<i>asnS</i>	Asparaginylyl-tRNA synthetase	0,513	50,0	5,03	x
1195	<i>asnS</i>	Asparaginylyl-tRNA synthetase	0,512	50,0	5,03	x
0338	<i>trpS</i>	Tryptophanyl-tRNA synthetase	0,508	38,3	5,45	
1332	<i>infB</i>	Translation initiation factor IF-2	0,508	90,9	9,23	
1249	<i>glyQ</i>	Glycyl-tRNA synthetase, alpha subunit	0,505	35,0	5,16	
0889	<i>aspS</i>	Aspartyl-tRNA synthetase	0,502	69,5	4,96	
1338	<i>proS</i>	Prolyl-tRNA synthetase	0,495	62,8	5,08	x
0256	<i>tyrS</i>	Tyrosyl-tRNA synthetase	0,486	47,6	5,55	x
1487	<i>pheS</i>	Phenylalanyl-tRNA synthetase, alpha subunit	0,483	39,7	5,35	x
1486	<i>pheT</i>	Phenylalanyl-tRNA synthetase, beta subunit	0,48	88,0	4,83	x
1342	<i>frr</i>	Ribosome recycling factor	0,474	20,7	6,03	
0701	<i>prfA</i>	Peptide chain release factor 1 (RF-1)	0,473	41,2	4,85	
Réponse au stress						
1313	<i>dnaK</i>	Chaperone protein DnaK (heat shock protein)	0,657	66,1	4,68	x*
1617	<i>groEL</i>	60 kDa chaperonin GroEL	0,59	57,3	4,86	x*

0624	<i>clpP</i>	ATP-dependent Clp protease proteolytic subunit	0,555	21,3	5,07	
0718	<i>uspI</i>	universal stress protein family	0,529	17,1	7,02	
0289		Small heat shock protein	0,501	15,8	4,77	
Divers						
0777	<i>tig</i>	Trigger factor	0,745	48,6	4,61	x
0853	<i>hIbA</i>	Histone-like DNA-binding protein	0,691	9,8	9,26	
1646	<i>hIbB</i>	Histone-like DNA-binding protein	0,687	9,7	9,3	
1533	<i>prtM</i>	Putative peptidylprolyl isomerase	0,665	33,1	9,87	
0995	<i>ppaC</i>	Manganese-dependent inorganic pyrophosphatase	0,611	34,3	4,76	x*
0763	<i>typA</i>	GTP-binding protein TypA	0,564	68,4	5,17	x
1612	<i>yajC</i>	Preprotein translocase, YajC subunit	0,553	14,4	4,84	
0105	<i>luxS</i>	S-ribosylhomocysteinase	0,552	17,9	5,09	x
1390		Putative kinase	0,521	59,9	4,38	x
1020	<i>fhs</i>	Formate-tetrahydrofolate ligase	0,517	61,4	5,43	x
2051		GTP-binding protein	0,502	40,4	4,86	
0319	<i>pncB</i>	Nicotinate phosphoribosyltransferase	0,494	56,1	5,4	x
0815		Inosine-uridine preferring nucleoside hydrolase	0,485	35,4	4,68	x
0607	<i>secA</i>	Preprotein translocase, SecA subunit	0,477	91,7	5,08	
0744	<i>ftsZ</i>	Cell division protein FtsZ	0,457	47,8	4,47	x
0368	<i>ftsH</i>	Cell division protein FtsH	0,451	80,3	5	
Hypothetical proteins						
1963		Hypothetical protein	0,779	29,1	10,94	
0163		Conserved hypothetical protein	0,767	4,9	11,45	
0349		Hypothetical protein	0,628	25,2	10,25	
0677		Conserved hypothetical protein	0,608	26,6	4,77	x
2081		Hypothetical protein	0,601	28,2	10,61	
0327		Hypothetical protein	0,595	10,3	6,28	
1604		Conserved hypothetical protein	0,577	12,1	3,49	
0568		Hypothetical protein	0,577	8,6	4,72	
0114		Hypothetical protein	0,569	26,1	5,49	x
1191		Hypothetical protein	0,551	11,9	9,3	
0165		Hypothetical protein	0,548	38,4	9,85	
0220		Conserved hypothetical protein	0,543	16,2	4,73	x
0125		Hypothetical protein	0,542	60,1	5,13	
1597		Conserved hypothetical protein	0,541	13,1	4,14	
2022		Conserved hypothetical protein	0,541	17,4	10,15	
0361		Conserved hypothetical protein	0,537	25,8	8,87	
1192		Hypothetical protein	0,521	8,1	9,3	
1354		Conserved hypothetical protein	0,515	8,3	10,24	
1523		Conserved hypothetical protein	0,513	23,0	4,62	x
0446		Hypothetical protein	0,513	5,8	9,57	
0097		Hypothetical protein	0,507	30,5	10,4	
0038		Hypothetical protein	0,501	102,3	10,31	
0592		Conserved hypothetical protein	0,5	14,8	3,87	
0871		Conserved hypothetical protein	0,499	39,1	4,75	x
2015		Conserved hypothetical protein	0,499	26,6	10,18	
1710		Conserved hypothetical membrane protein	0,497	20,1	8,89	
0161		Conserved hypothetical protein	0,49	20,9	5,96	x
1560		Hypothetical protein	0,484	33,1	4,93	
1404		Hypothetical protein	0,481	19,5	4,98	
2020		Conserved hypothetical protein	0,466	17,3	10,21	

A l'intérieur des catégories fonctionnelles, les gènes sont triés selon leur CAI. Les produits des gènes dont le pI n'est pas compatible avec une détection sur le gel de référence (p>I compris entre 4 et 7) sont grisés. x : identifié sur gel de référence. * identifié lors de l'étude en 2DE de l'adaptation au stress acide (voir partie II).

Les 209 gènes de la classe III

n° ldb	gène	produit	GC%	GC3%
Biosynthèse des EPS				
1695	<i>rfbA</i>	Putative glucose-1-phosphate thymidyltransferase	0,383	0,379
1942	<i>epsIIJ</i>	EpsIIJ, Hypothetical membrane protein	0,367	0,363
1943	<i>epsII-I</i>	EpsII-I, Putative glycosyltransferase	0,347	0,381
1944	<i>epsIIH</i>	EpsIIH, Putative glycosyltransferase	0,386	0,349
1945	<i>epsIIG</i>	EpsIIG, Putative glycosyltransferase	0,404	0,379
1998	<i>epsIM</i>	EpsIM, Putative glycosyltransferase	0,371	0,342
2001	<i>epsIK</i>	EpsIK, Putative polysaccharide repeat unit transport protein	0,4	0,426
2003	<i>epsI-I</i>	EpsI-I, Hypothetical membrane protein	0,342	0,281
2004	<i>epsIH</i>	EpsIH, Putative glycosyltransferase	0,328	0,307
2005	<i>epsIG</i>	EpsIG, Putative glycosyltransferase	0,349	0,335
2006	<i>epsIF</i>	EpsIF, Hypothetical protein	0,366	0,351
2008	<i>epsID</i>	EpsID, Hypothetical protein	0,462	0,492
Transport				
14		ABC transporter, ATP-binding/permease protein	0,369	0,387
154		ABC transporter, ATP-binding protein	0,433	0,49
156		ABC transporter, permease protein	0,447	0,512
251		Amino Acid ABC transporter, permease protein	0,359	0,306
252		Amino acid ABC transporter, permease protein	0,362	0,296
253		Amino acid ABC transporter, ATP-binding protein	0,36	0,272
254		Amino acid ABC transporter, substrate binding protein	0,368	0,294
302		ABC transporter, ATP-binding/permease protein	0,443	0,507
330		ABC transporter, permease protein	0,397	0,429
331		ABC transporter, ATP-binding protein	0,411	0,39
1239		Cation transporting P-type ATPase (probable copper transporter)	0,429	0,406
1699		Permease	0,426	0,49
1828		ABC transporter, ATP-binding protein	0,446	0,5
2094		ABC transporter, ATP-binding/permease protein	0,376	0,354
2178	<i>potB2</i>	Spermidine/putrescine ABC transporter, permease protein	0,377	0,362
2179	<i>potC2</i>	Spermidine/putrescine ABC transporter, permease protein	0,37	0,351
2180	<i>potA2</i>	Spermidine/putrescine ABC transporter, ATP-binding protein	0,378	0,331
2181	<i>potD2</i>	Spermidine/putrescine ABC transporter, substrate binding protein	0,393	0,371
2208		ABC transporter, ATP-binding protein	0,449	0,531

Transport d'électrons et ATP synthase

707	<i>atpF</i>	H ⁺ transporting ATPase / ATP synthase, B subunit	0,469	0,494
708	<i>atpH</i>	H ⁺ transporting ATPase / ATP synthase, delta subunit	0,459	0,5
1095		Putative fumarate reductase (flavoprotein)	0,426	0,37
1112		Putative fumarate reductase (flavoprotein)	0,426	0,37
Métabolisme des acides nucléiques				
2110	<i>carA1, pyrAA1</i>	Carbamoyl-phosphate synthase, small chain	0,424	0,374
2111	<i>pyrC1</i>	Dihydroorotase	0,406	0,311
2112	<i>pyrB1</i>	Aspartate carbamoyltransferase	0,396	0,371
2113	<i>pyrR1</i>	Pyrimidine operon regulatory protein / Uracil phosphoribosyltransferase	0,403	0,294
2114	<i>pyrD1</i>	Dihydroorotate dehydrogenase	0,415	0,326
2182	<i>adeC</i>	Adenine deaminase	0,381	0,296
Biosynthèse de vitamines				
246	<i>folB</i>	Dihydroneopterin aldolase	0,402	0,405
247	<i>folKE</i>	2-Amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase / GTP cyclohydrolase I	0,411	0,392
248	<i>folC2</i>	Folylpolyglutamate synthase	0,399	0,362
249	<i>folP</i>	Dihydropteroate synthase	0,39	0,347
Maturation et modification des protéines				
265		Aminoacylase	0,388	0,304
1086		Putative phosphotyrosine protein phosphatase	0,484	0,538
1121		Putative phosphotyrosine protein phosphatase	0,484	0,538
1165		Acyltransferase	0,457	0,505
1189	<i>prtB</i>	Proteinase precursor	0,476	0,542
Restriction et modification de l'ADN				
474		Putative restriction endonuclease	0,434	0,414
1230	<i>mod</i>	Type III restriction-modification system methyltransferase	0,377	0,301
1162		Putative exonuclease	0,455	0,548
1053	<i>hsdS1</i>	Type I restriction-modification system, specificity subunit	0,371	0,316
1054		Integrase/recombinase	0,381	0,377
1055	<i>hsdS2</i>	Type I restriction-modification system, specificity subunit	0,381	0,328
Régulation de la transcription				
159		Putative transcriptional regulator	0,402	0,448
303		Transcriptional regulator (AraC family)	0,428	0,496
1793		Putative transcriptional regulator (Xre family)	0,433	0,466
Transposons et IS				
83		Hypothetical protein	0,429	0,407
100		Transposase	0,459	0,501
326		Putative Transposase	0,436	0,486
803		Transposase	0,437	0,486

835		Transposase	0,438	0,477
919		Transposase	0,438	0,486
1206		Transposase	0,439	0,479
1320		Transposase	0,442	0,491
1393		Transposase	0,455	0,497
1466		Transposase	0,438	0,477
1752		Transposase	0,44	0,482
1782		Transposase	0,39	0,344
1803		Transposase	0,441	0,491
1874		Transposase	0,44	0,482
1907		Transposase	0,462	0,518
1934		Transposase	0,436	0,477
1979		Transposase	0,438	0,477
2033		Putative transposase	0,435	0,484
2153		Transposase	0,44	0,482
Divers				
416	<i>secY1</i>	Preprotein translocase, SecY subunit	0,448	0,492
1325	<i>cysK2</i>	Cysteine synthase	0,456	0,422
1461	<i>cfa</i>	Cyclopropane-fatty-acyl-phospholipid synthase	0,374	0,336
1490		Putative acylphosphatase	0,498	0,544
1502	<i>infC</i>	Translation initiation factor IF-3	0,473	0,509
1775		Ornithine decarboxylase	0,354	0,227

Protéines hypothétiques ou de fonction inconnue*123 protéines non listées*

A l'intérieur des catégories fonctionnelles, les gènes sont triés selon leur n° lb. GC% : pourcentage en G-C du gène. GC3% : pourcentage en G-C de la troisième base des codons du gène. Les crochets représentent les gènes groupés dans le génome.

Annexe 3 : Voies de biosynthèse des acides aminés

Acide aminé	Enzyme	EC number	présent	frag.	absent	Voies
Alanine						4
	Valine-pyruvate transaminase	2.6.1.66			x	1
	Alanine aminotransferase	2.6.1.2			x	2
	Aspartate decarboxylase	4.1.1.12			x	3
	Alanine dehydrogenase	1.4.1.1			x	4
Arginine						4
	Glutamate 5-kinase	2.7.2.11		x		1
	Glutamate 5-semialdehyde dehydrogenase	1.2.41			x	1
	Ornithine aminotransferase	2.6.1.13			x	1
	Aminoacid N-acetyltransferase	2.3.1.1			x	2/3/4
	Acetylglutamate kinase	2.7.2.8			x	2/3/4
	N-acetyl-g-glutamyl-phosphate reductase	1.2.1.38			x	2/3/4
	Acetylmithine aminotransferase	2.6.1.11			x	2/3/4
	Acetylmithine deacetylase	3.5.1.16			x	2
	Aminoacylase	3.5.1.14	x			3
	Glutamate acetyltransferase	2.3.1.35			x	4
	Carbamoylphosphate synthase	6.3.5.5	x	x		1/2/3/4
	Ornithine carbamoyltransferase	2.1.3.3			x	1/2/3/4
	Argininosuccinate synthase	6.3.4.5			x	1/2/3/4
	Argininosuccinate lyase	4.3.2.1		x		1/2/3/4
Asparagine	Voies 1 et 2 potentiellement fonctionnelles					2
	Aspartate-ammonia ligase	6.3.1.1	x			1
	Asparagine synthetase	6.3.5.4	x			2
Aspartate	Voies 1 et 2 potentiellement fonctionnelles					3
	Aspartate ammonia-lyase	4.3.1.1	x			1
	Asparagine synthetase	6.3.5.4	x			2
	Aspartate aminotransférase	2.6.1.1			x	3
Cystéine						2
	Serine O-acetyltransferase	2.3.1.30		x		1
	Cystéine synthase	2.5.1.47	x			1
	cystathionine beta-synthase	4.2.1.22			x	2
	Cystathionine gamma synthase	4.4.1.1		x		2
Glutamate	Voie 4 potentiellement fonctionnelle*					4
	Glutamate dehydrogenase	1.4.1.2/3/4			x	1
	Aspartate aminotransferase	2.6.1.1			x	2
	Glutamate synthase	1.4.7.1&1.4.1.13/14			x	3
	Glutamine synthetase	6.3.1.2	x			4
Glutamine	Voie 1 potentiellement fonctionnelle*					1
	Glutamine synthetase	6.3.1.2	x			1

Acide aminé	Enzyme	EC number	présent	frag.	absent	Voies
Glycine						2
	Glycine hydroxymethyltransferase	2.1.2.1			x	1
	Threonine aldolase	4.1.2.5			x	2
Histidine						1
	Ribokinase	2.7.1.15	x			1
	Ribose-P pyrophosphokinase	2.7.6.1	x			1
	ATP Phosphoribosyltransferase	2.4.2.17			x	1
	Phosphoribosyl-ATP pyrophosphatase	3.6.1.31			x	1
	Phosphoribosyl-AMP cyclohydrolase	3.5.4.19			x	1
	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino) methylideneamino] imidazole-4-carboxamide isomerase	5.3.1.16			x	1
	Pentosyl ribosyltransferase	2.4.2.-			x	1
	imidazoleglycerol-phosphate dehydratase	4.2.1.19			x	1
	Histidinol phosphate aminotransferase	2.6.1.9			x	1
	Histidinol phosphatase	3.1.3.15			x	1
	Histidinol dehydrogenase	1.1.1.23			x	1
Isoleucine						1
	threonine ammonia-lyase	4.3.1.19/4.2.1.16			x	1
	Acetolactate synthase	4.1.3.18			x	1
	Ketol-acid reductoisomerase	1.1.1.86			x	1
	dihydroxy-acid dehydratase	4.2.1.9			x	1
	Branched chain amino acid aminotransferase	2.6.1.42	x			1
Leucine						2
	Acetolactate synthase	4.1.3.18			x	1/2
	2-acetolactate mutase	5.4.99.3			x	1/2
	Ketol-acid reductoisomerase	1.1.1.86			x	1/2
	Dihydroxy-acid dehydratase	4.2.1.9			x	1/2
	2-isopropylmalate synthase	4.1.3.12			x	1/2
	3-isopropylmalate dehydratase	4.2.1.33			x	1/2
	3-isopropylmalate dehydrogenase	1.1.1.85			x	1/2
	Leucine dehydrogenase	1.4.1.9			x	1
	Branched chain amino acid aminotransferase	2.6.1.42	x			2
Lysine						1
	Aspartate kinase	2.7.2.4	x			1
	Aspartate-semialdehyde dehydrogenase	1.2.1.11	x			1
	dihydrodipicolinate synthase	4.2.1.52			x	1
	dihydrodipicolinate reductase	1.3.1.26			x	1
	tetrahydrodipicolinate N-succinyltransferase	2.3.1.117			x	1
	succinyldiaminopimelate aminotransferase	2.6.1.17			x	1
	succinyl-diaminopimelate desuccinylase	3.5.1.18			x	1
	diaminopimelate epimerase	5.1.1.7			x	1
	Diaminopimelate decarboxylase	4.1.1.20		x		1
Methionine						2
	Aspartate kinase	2.7.2.4	x			1/2
	Aspartate-semialdehyde dehydrogenase	1.2.1.11	x			1/2
	Homoserine dehydrogenase	1.1.1.3	x			1/2
	Homoserine O-succinyltransferase	2.3.1.46		x		1/2
	Cystathionine gamma synthase	4.2.99.9		x		1/2
	Cystathionine beta-lyase	4.4.1.8	x			1
	Homocysteine S-methyltransferase	2.1.1.10	x			1/2

Acide aminé	Enzyme	EC number	présent	frag.	absent	Voies
Phénylalanine						3
	Phosphoketolase	4.1.2.9	x			1/2/3
	2-Dehydro-3-deoxyphosphoheptanoate aldolase	4.1.2.15			x	1/2/3
	3-dehydroquinone synthase	4.2.3.4			x	1/2/3
	3-dehydroquinone dehydratase	4.2.1.10			x	1/2/3
	Shikimate 5-dehydrogenase	1.1.1.25			x	1/2/3
	Shikimate kinase	2.7.1.71		x		1/2/3
	3-phosphoshikimate 1-carboxyvinyltransferase	2.5.1.19		x		1/2/3
	Chorismate synthase	4.2.3.5		x		1/2/3
	Chorismate mutase	5.4.99.5	x			1/2/3
	Aromatic amino acid transferase	2.6.1.57	x			1
	Carboxycyclohexadienyl dehydratase	4.2.1.91			x	1
	Prephenate dehydratase	4.2.1.51			x	2/3
	Phénylalanine dehydrogenase	1.4.1.20			x	2
	Aromatic amino acid transferase	2.6.1.57	x			3
Proline						5
	Glutamate 5-kinase	2.7.2.11		x		1/2
	Glutamate 5-semialdehyde dehydrogenase	1.2.41			x	1/2
	Pyroline-5-carboxylate reductase	1.5.1.2			x	1
	Ornithine aminotransferase	2.6.1.13			x	2
	Aminoacid N-acetyltransferase	2.3.1.1			x	3/4/5
	Acetylglutamate kinase	2.7.2.8			x	3/4/5
	N-acetyl-g-glutamyl-phosphate reductase	1.2.1.38			x	3/4/5
	Acetylornithine aminotransferase	2.6.1.11			x	3/4/5
	Acetylornithine deacetylase	3.5.1.16			x	3
	Aminoacylase	3.5.1.14	x			4
	Glutamate acetyltransferase	2.3.1.35			x	5
	Ornithine cyclodeaminase	4.3.1.12			x	2/3/4/5
Serine						3
	Phosphoglycerate dehydrogenase	1.1.1.95		x		1
	Phosphoserine aminotransferase	2.6.2.51		x		1
	Phosphoserine phosphatase	3.1.3.3			x	1
	Phosphoglycerate phosphatase	3.1.3.20			x	2
	Glycerate dehydrogenase	1.1.1.29			x	2
	Serine-pyruvate aminotransferase	2.6.1.51			x	2
	Serine dehydratase	4.3.1.17			x	3
Thréonine	Voie 1 potentiellement fonctionnelle					2
	Aspartate kinase	2.7.2.4	x			1
	Aspartate-semialdehyde dehydrogenase	1.2.1.11	x			1
	Homoserine dehydrogenase	1.1.1.3	x			1
	Homoserine kinase	2.7.1.39	x			1
	Threonine synthase	4.2.3.1	x			1
	Threonine aldolase	4.1.2.5			x	2

Acide aminé	Enzyme	EC number	présent	frag.	absent	Voies
Tryptophane						1
	Phosphoketolase	4.1.2.9	x			1
	2-Dehydro-3-deoxyphosphoheptanoate aldolase	4.1.2.15			x	1
	3-dehydroquinatase synthase	4.2.3.4			x	1
	3-dehydroquinatase déshydratase	4.2.1.10			x	1
	Shikimate 5-déshydrogénase	1.1.1.25			x	1
	Shikimate kinase	2.7.1.71		x		1
	3-phosphoshikimate 1-carboxyvinyltransférase	2.5.1.19		x		1
	Chorismate synthase	4.2.3.5		x		1
	Anthranilate synthase	4.1.3.27			x	1
	Anthranilate phosphoribosyltransférase	2.4.2.18			x	1
	Phosphoribosylanthranilate isomérase	5.3.1.24			x	1
	Indole-3-glycérol-phosphate synthase	4.1.1.48			x	1
	Tryptophan synthase	4.2.1.20			x	1
Tyrosine						3
	Phosphoketolase	4.1.2.9	x			1/2/3
	2-Dehydro-3-deoxyphosphoheptanoate aldolase	4.1.2.15			x	1/2/3
	3-dehydroquinatase synthase	4.2.3.4			x	1/2/3
	3-dehydroquinatase déshydratase	4.2.1.10			x	1/2/3
	Shikimate 5-déshydrogénase	1.1.1.25			x	1/2/3
	Shikimate kinase	2.7.1.71		x		1/2/3
	3-phosphoshikimate 1-carboxyvinyltransférase	2.5.1.19		x		1/2/3
	Chorismate synthase	4.2.3.5		x		1/2/3
	Chorismate mutase	5.4.99.5	x			1/2/3
	Aromatic amino acid transférase	2.6.1.57	x			1
	Cyclohexadienyl déshydrogénase	1.3.1.43			x	1
	Prephenate déshydrogénase	1.3.1.12			x	2/3
	Phénylalanine déshydrogénase	1.4.1.20			x	2
	Aromatic amino acid transférase	2.6.1.57	x			3
Valine						1
	Acéto lactate synthase	4.1.3.18			x	1
	2-acéto lactate mutase	5.4.99.3			x	1
	Kéto lactate réductoisomérase	1.1.1.86			x	1
	Dihydroxy-acide déshydratase	4.2.1.9			x	1

Récapitulatif des enzymes des voies de biosynthèse des acides aminés.

Pour chaque acide aminé, les enzymes permettant sa biosynthèse sont indiquées avec leur numéro EC. Pour chaque enzyme, il est indiqué si son gène est présent intégralement dans le génome (présent), s'il est présent sous forme fragmentée (frag.) ou s'il est absent (absent). Le nombre de voies de biosynthèse pour chaque acide aminé (source : <http://www.genome.ad.jp/kegg/>) est indiqué (ligne grise) et pour chaque enzyme, il est indiqué à quelle voie elle appartient (les voies sont numérotées arbitrairement).

* Voie qui, pour être active, nécessite la présence d'un acide aminé non synthétisable par ATCC11842 : pour la biosynthèse de glutamate, besoin de glutamine et pour la biosynthèse de glutamine, besoin de glutamate

Annexe 4 : Conservation des régulateurs transcriptionnels

Spécifiques à *L. bulgaricus*

Ldb0048	Putative transcriptional regulator
Ldb0159	Putative transcriptional regulator
Ldb0153	Putative transcriptional regulator (AraC family)
Ldb0493	Transcriptional regulator (AsnC family)
Ldb0308	Putative transcriptional regulator (Crp family)
Ldb1182	Transcriptional regulator (DeoR family)
Ldb2068	Putative transcriptional regulator (DeoR family)
Ldb1715	Transcriptional regulator (GntR family)
Ldb0115	Transcriptional regulator (LysR family)
Ldb0124	Transcriptional regulator (LysR family)
Ldb0808	Transcriptional regulator (LysR family)
Ldb0856	Transcriptional regulator (LysR family)
Ldb1290	Transcriptional regulator (LysR family)
Ldb1683	Transcriptional regulator (LysR family)
Ldb1984	Transcriptional regulator (LysR family)
Ldb1953	EpsIIA, Transcriptional regulator of eps operon (LytR family)
Ldb0314	Putative transcriptional regulator (MarR family)
Ldb1075	Putative transcriptional regulator of multidrug efflux pump gene (MerR family)
Ldb1132	Putative transcriptional regulator of multidrug efflux pump gene (MerR family)
Ldb2141	Transcriptional regulator (TetR family)
Ldb1793	Putative transcriptional regulator (Xre family)
Ldb1958	Putative transcriptional regulator (Xre family)

Spécifiques à *L. acidophilus*

LBA0228	transcriptional regulator
LBA1461	putative regulator
LBA0593	putative transcriptional regulator
LBA1691	putative transcriptional regulator
LBA1701	melibiose operon regulatory protein, AraC family
LBA0355	transcriptional regulator, AraC family
LBA1700	transcriptional regulator, AraC family
LBA1899	putative transcriptional regulator, ArsR family
LBA0875	transcriptional regulator, GntR family
LBA0500	msm operon repressor, LacI family
LBA1074	malolactic regulator, LysR family
LBA0065	transcriptional regulator, LysR family
LBA0458	transcriptional regulator, LysR family
LBA1054	transcriptional regulator, LysR family
LBA1410	transcriptional regulator, LysR family
LBA1136	transcriptional regulator, MarR family
LBA1293	putative transcriptional regulator, MarR family
LBA1021	putative transcriptional regulator, MerR family
LBA0228	transcriptional regulator, NagC family
LBA1934	transcriptional regulator, Rgg family
LBA0226	transcriptional regulator, RpiR family
LBA1172	transcriptional regulator, RpiR family
LBA1577	transcriptional regulator, RpiR family
LBA0554	transcriptional regulator, TetR family
LBA0085	transcriptional regulator, Xre family
LBA1788	transcriptional regulator, Xre family
LBA0244	putative transcriptional regulator, Xre family
LBA1863	putative transcriptional regulator, Xre family

Spécifiques à *L. johnsonii*

LJ0075	transcriptional regulator, AraC family
LJ0185	transcriptional regulator, GntR family
LJ0116	transcriptional regulator, GntR family
LJ0126	transcriptional regulator, GntR family
LJ0744	transcriptional regulator, LacI family
LJ0050	transcriptional regulator, LysR family
LJ1114	transcriptional regulator, LysR family
LJ1178	transcriptional regulator, LysR family
LJ0396	transcriptional regulator, LytR family
LJ0724	transcriptional regulator, MarR family
LJ1048	putative transcriptional regulator, MerR family
LJ1796	putative transcriptional regulator, Rgg family
LJ0528	transcriptional regulator, RpiR family
LJ0144	transcriptional regulator, RpiR family
LJ0242	transcriptional regulator, TetR family
LJ0880	transcriptional regulator, TetR family
LJ0105	putative transcriptional regulator, TetR family
LJ1054	transcriptional regulator, Xre family
LJ1831	transcriptional regulator, Xre family

Communs à *L. acidophilus* et *L. johnsonii*

LBA0237	LJ0259	transcriptional regulator raffinose operon, AraC family
LBA1774	LJ0576	transcriptional regulator, ArsR family
LBA0599	LJ0802	transcriptional regulator, DeoR family
LBA1013	LJ0759	trehalose operon transcriptional repressor, GntR family
LBA0393	LJ1662	transcriptional regulator, GntR family
LBA0882	LJ0189	transcriptional regulator, GntR family
LBA1488	LJ1265	putative ribose operon repressor, LacI family
LBA1472	LJ1260	ribose operon repressor, LacI family
LBA0399	LJ0517	sucrose operon repressor, LacI family
LBA1465	LJ0856	transcription repressor of beta-galactosidase gene, LacI family
LBA1874	LJ0210	transcriptional regulator, LacI family
LBA1024	LJ1118	putative transcriptional regulator, LysR family
LBA1124	LJ1172	putative transcriptional regulator, LysR family
LBA0596	LJ1156	transcriptional regulator, MarR family
LBA1663	LJ0661	transcriptional regulator, MarR family
LBA0573	LJ0749	putative transcriptional regulator, MarR family
LBA1955	LJ1828	transcription regulator, MerR family
LBA1444	LJ1575	transcriptional regulator, MerR family
LBA1368	LJ0734	transcriptional xylose repressor, NagC family
LBA1367	LJ0735	transcriptional regulator, NagC family
LBA0886	LJ0195	sugar kinase -putative transcriptional regulator, NagC family
LBA1688	LJ0637	transcriptional regulator, RpiR family
LBA1702	LJ1817	transcriptional regulator, RpiR family
LBA0607	LJ0637	putative transcriptional regulator, RpiR family
LBA1614	LJ0238	putative transcriptional regulator, TetR family
LBA0663	LJ0837	putative transcriptional regulator, Xre family
LBA0896	LJ0294	putative transcriptional regulator, Xre family

Communs à *L. bulgaricus* et *L. johnsonii*

Ldb0303	LJ0708	Transcriptional regulator (AraC family)
Ldb1307	LJ0376	Transcriptional regulator (ArsR family)

Communs à <i>L. bulgaricus</i> et <i>L. acidophilus</i>			
Ldb0183	LBA1969		Transcriptional regulator (Crp family)
Ldb2082	LBA0925		Transcriptional regulator (DeoR family)
Ldb0181	LBA0132		Putative transcriptional regulator (TetR family)
Ldb1638	LBA0563		Pyrimidine operon regulatory protein
Communs à <i>L. bulgaricus</i>, <i>L. acidophilus</i> et <i>L. johnsonii</i>			
Ldb0634	LBA0697	LJ0871	Central glycolytic genes regulator
Ldb1315	LBA1249	LJ1481	Heat-inducible transcription repressor HrcA
Ldb1356	LBA1280	LJ1508	Transcriptional repressor of the SOS regulon (LexA)
Ldb0579	LBA0642	LJ0819	Negative regulator of competence MecA
Ldb0482	LBA0544	LJ1693	Transcriptional regulator (Crp family)
Ldb2042	LBA1779	LJ0570	Transcriptional regulator (DeoR family)
Ldb0449	LBA0516	LJ1739	Transcriptional regulator (GntR family)
Ldb0533	LBA0592	LJ0800	Transcriptional regulator (GntR family)
Ldb1689	LBA1189	LJ1302	Transcriptional regulator (GntR family)
Ldb1300	LBA1233	LJ1466	Transcriptional regulator (LysR family)
Ldb1593	LBA0431	LJ0491	Catabolite control protein A (CcpA), LacI family
Ldb2011	LBA1737	LJ1021	EpsIA, Transcriptional regulator of eps operon (LytR family)
Ldb0231	LBA0188	LJ0172	Putative transcriptional regulator (LytR family)
Ldb2032	LBA1762	LJ0582	Putative transcriptional regulator (LytR family)
Ldb0658	LBA1963	LJ1835	Putative transcriptional regulator (MarR family)
Ldb1488	LBA1520	LJ1626	Putative transcriptional regulator (MarR family)
Ldb1892	LBA1638	LJ0666	Putative transcriptional regulator (TetR family)
Ldb2062	LBA1835	LJ0547	Putative transcriptional regulator (TetR family)
Ldb2071	LBA1840	LJ0539	Putative transcriptional regulator (TetR family)
Ldb1886	LBA1630	LJ0672	Putative transcriptional regulator (Xre family)
Ldb2113	LBA1383	LJ1280	Pyrimidine operon regulatory protein

Pour chaque gène, les « locus_tag » sont donnés. Ldbxxxx sont les gènes de *L. bulgaricus*, LBAxxxx ceux de *L. acidophilus* et LJxxxx ceux de *L. johnsonii*. Les résultats ont été obtenus par analyse BSR (Blast Score Ratio ; Rasko *et al.*, 2005), avec un seuil de 0,4.

Annexe 5 : Bilan de la recherche des mécanismes de résistance au stress

Tableau 1 : Résistance au stress oxydant

Gène	L. bulg	n° de gène	fonction
<i>ahp</i>			peroxiredoxin
<i>ahpC</i>			alkyl hydroperoxide reductase(small subunit)
<i>ahpF</i>			alkyl hydroperoxide reductase(large subunit)
<i>dps</i>			stress- and starvation-induced gene
<i>fni</i>	x	<i>ldb0996</i>	Isopentenyl-diphosphate delta-isomerase
<i>fur</i>			transcriptional regulator
<i>glpQ</i>	x	<i>ldb1925</i>	Glycerophosphoryl diester phosphodiesterase
<i>hemA</i>			Glutamyl tRNA reductase
<i>hemB</i>			delta-aminolevulinic acid dehydratase
<i>hemC</i>			porphobilinogen deaminase
<i>hemD</i>			uroporphyrinogen III cosynthase
<i>hemX</i>			negative effector of the concentration of HemA
<i>hslO</i>	x	<i>ldb0369</i>	redox regulated chaperonin
<i>iscU</i>	x	<i>ldb1818</i>	NifU homolog involved in Fe-S cluster formation
<i>ispA</i>	x	<i>ldb1422</i>	Geranylgeranyl pyrophosphate synthase
<i>ispB</i>			octaprenyl-diphosphate synthase
<i>katE/A/B/G</i>			catalase
<i>msrA</i>	x	<i>ldb0673</i>	Peptide methionine sulfoxide reductase
<i>msrB</i>	x	<i>ldb1971</i>	Protein-methionine-S-oxide reductase
<i>odc</i>	x	<i>ldb1775/0547</i>	ornithine decarboxylase
<i>oxyR</i>			transcriptional regulator
<i>pot</i>	x	<i>ldb2180/0647</i>	spermidine/putrescine ABC transporter
<i>pox</i>	x	<i>ldb2213</i>	pyruvate oxidase
<i>sodA</i>			superoxide dismutase
<i>soxR</i>			transcriptional regulator
<i>sufB/D</i>	x	<i>ldb1817/1820</i>	Conserved protein involved in Fe/S cluster assembly
<i>sufC</i>	x	<i>ldb1821</i>	ATP-binding protein involved in Fe/S cluster assembly
<i>sufS</i>	x	<i>ldb1819</i>	Cysteine desulfurase / Selenocysteine lyase
<i>trxA</i>	x	<i>ldb1524/ldb1602</i>	thioredoxin
<i>trxB</i>	x	<i>ldb0613/ldb1586</i>	thioredoxin reductase
<i>zur</i>			Fe(II)/Zn(II) uptake regulation proteins
			NADH:H ₂ O ₂ oxidase
			NADH:H ₂ O oxidase
			NADH:peroxidase
			oxidase glycerophosphate
			L-Cys uptake system
			glutathione peroxidase
			phytoene/squalene synthetase
			phytoene DH
			lycopene cyclase: FAD/NAD(P) binding protein
			lysophospholipase
			carboxylesterase
	x	<i>ldb1885/ldb0191</i>	esterase/lipase
			arginine decarboxylase
			spermidine synthase
			NRAMP family transporter
	x	<i>ldb2037-8</i>	ABC-type Mn/Zn transport systems (permease)
	x	<i>ldb0675</i>	ABC-type metal ion transport system
			Mn-dependent transcriptional regulator
			Lysine/ornithine N-monoxygenase
			Siderophore biosynthesis protein
			Siderophore synthetase component
			Ferritin-like protein
			outer membrane receptor, mostly Fe transport
			outer membrane receptor for ferrienterochelin and colicins
			ABC-type Fe(III)-siderophore transport system
			siderophore interacting protein
			Fe(II) transport system protein A/B
			NO synthase

Tableau 2 : Résistance au stress osmotique

gène	<i>L. bulg.</i>	n° de gène	fonction
<i>aqpZ</i>			aquaporin
<i>befT</i>			choline transporter
<i>busA</i>			glycine betaine uptake system
<i>busR</i>			transcriptional regulator
<i>des</i>			fatty acid desaturase
<i>ftsH</i>	x	<i>ldb0368</i>	Cell division protein FtsH
<i>gbsA</i>			glycine betaine aldehyde dehydrogenase
<i>gbsB</i>			alcohol dehydrogenase
<i>glnPQ</i>	x	<i>ldb2198-2199</i>	Glutamine ABC transporter
<i>glnQHMP</i>	x	<i>ldb517-521</i>	Glutamine ABC transporter
<i>htrA</i>	x	<i>ldb0140</i>	HtrA-like serine protease
<i>kdp</i>			High-affinity potassium transport
<i>kup</i>	x	<i>Ldb0219</i>	K ⁺ uptake permease
<i>mscL</i>	x	<i>ldb0318</i>	Large-conductance mechanosensitive channel
<i>mscS</i>	x	<i>ldb1602</i>	Mechanosensitive ion channel
<i>ompC</i>			Outer membrane protein 1b
<i>ompF</i>			Outer membrane protein 1a
<i>opuA</i>			Glycine betaine ABC transporter
<i>opuC</i>	x	<i>ldb1180/1181</i>	Glycine betaine ABC transporter
<i>opuD</i>			glycine betaine transporter
<i>opuE</i>			proline transporter
<i>proA</i>			gamma-glutamyl phosphate reductase
<i>proB</i>	fgt	<i>ldb0830-2</i>	glutamate 5 kinase
<i>proP</i>			betaine, proline, ectoine transporter
<i>putP</i>			Proline/Na ⁺ , Li ⁺ symport protein
<i>qacT</i>			glycine betaine uptake system
<i>trk</i>			potassium uptake

Tableau 3 : Résistance au stress acide

Gène	<i>L. bulg.</i>	N°lb	Function
<i>clpP</i>	x	<i>ldb0624</i>	ATP-dependent Clp protease proteolytic subunit
<i>htrA</i>	x	<i>ldb0140</i>	HtrA-like serine protease
<i>recA</i>	x	<i>ldb0599</i>	Recombinase A
<i>uvrA</i>	x	<i>ldb0615</i>	Excinuclease ABC, subunit A
<i>fabM</i>			Fatty acid desaturase
<i>dagK</i>			Diacylglycerol kinase
<i>ffh</i>	x	<i>ldb1372</i>	Signal recognition particle protein Ffh
<i>dltC</i>	x	<i>ldb2146</i>	D-alanyl carrier protein
<i>atpA-H</i>	x	<i>ldb0705-ldb0712</i>	H ⁺ transporting ATPase / ATP synthase
<i>gadCD</i>			Glutamate decarboxylase
<i>adc</i>			Aspartate decarboxylase
<i>hdc</i>			Histidine decarboxylase
<i>odc</i>	x	<i>ldb0547, ldb1775</i>	Ornithine décarboxylase
<i>mlf</i>			Malolactic fermentation
<i>citC-G, M, P</i>			Citrolactic fermentation
<i>arcABC</i>			Arginine deiminase
<i>ureA-G</i>			Urease

Les gènes présents dans le génome de *L. bulgaricus* sont marqués d'une croix dans la colonne *L. bulg.*
fgt : fragment, gène présent sous forme de pseudogène.

Les données viennent d'une synthèse bibliographique. Pour revue, voir Ghosal *et al.*, 2005 pour le stress oxydant et Sleator et Hill, 2001 pour le stress osmotique.

Résultats – Partie II
Etude de l'adaptation à l'acidité

Après cette première partie de ma thèse consacrée à l'analyse des données génomiques, j'ai utilisé des approches de physiologie microbienne et de biologie moléculaire pour aborder l'étude de la tolérance au stress acide de *L. bulgaricus*. En effet, nos analyses *in silico* semblent indiquer que cette bactérie lactique présente peu de gènes déjà décrits comme importants pour la réponse au stress acide chez les BL. Nous avons donc cherché à identifier des membres du stimulon acide par différentes approches.

1. Avant-propos

L. bulgaricus, comme toutes les BL, produit de l'acide lactique au cours de sa croissance et acidifie son milieu. La résistance au stress induite par l'acidification progressive du milieu est importante pour la survie de la bactérie dans le yaourt, survie rendue obligatoire par la législation en vigueur en France. D'autre part, le maintien d'un métabolisme à bas pH est responsable du phénomène de post-acidification. En effet, alors que sa croissance est inhibée par le pH, *L. bulgaricus* maintient une activité métabolique et continue de produire de l'acide lactique, ce qui produit une acidification accrue du yaourt au cours de la conservation du produit. La post-acidification est difficilement maîtrisable et induit une altération du produit au cours de la conservation. La compréhension des mécanismes de tolérance et d'adaptation au stress acide est donc importante pour mieux maîtriser l'ensemble du procédé industriel.

Les connaissances sur les mécanismes de résistance au stress acide chez *L. bulgaricus* ne sont pas très étendues. Une étude menée au laboratoire sur la souche Vi1010 (Lim *et al.*, 2000), a montré que *L. bulgaricus* est capable d'induire une ATR. Cependant, en l'absence de séquence génomique disponible, l'analyse en électrophorèse bidimensionnelle n'a permis d'identifier que trois protéines induites, les protéines chaperons GroES, GroEL et DnaK. Une autre étude portant sur l'ATR de la souche Vi-1010 induite par acidification à l'aide de HCl a permis de mettre en évidence 50 protéines induites par l'ATR, dont 21 ont été identifiées par séquençage N-terminal (Guillouard *et al.*, 2004). Néanmoins, plus de la moitié des protéines induites n'ont pu être identifiées.

Des études préliminaires menées sur la souche ATCC11842 (souche séquencée) ont montré qu'elle est capable de mettre en place une ATR. Cependant, l'analyse du génome n'apporte pas de pistes claires sur les mécanismes mis en jeu lors de cette adaptation. Pour identifier les fonctions nécessaires au développement de l'ATR ou de la résistance chez *L. bulgaricus*, nous ne disposons pas de méthode d'inactivation aléatoire de gènes qui aurait pu nous permettre

d'obtenir et de caractériser des mutants plus résistants ou plus sensibles au stress acide comme cela a été fait chez d'autres BL, par exemple chez *L. lactis* (Rallu *et al.*, 2000).

Dans cette partie de ma thèse, j'ai cherché à caractériser l'ATR à trois niveaux complémentaires : (1) en mesurant la variation de l'expression des régulateurs d'ATCC 11842 et (2) de celle de transporteurs d'ions susceptibles de contribuer à la force protomotrice et (3) par l'approche classique d'électrophorèse bi-dimensionnelle pour repérer les protéines cytoplasmiques sur- ou sous-exprimées. Les études (1) et (2) ont été réalisées sur de cultures en MPL, milieu chimiquement défini permettant la croissance de *L. bulgaricus*. Du fait d'une certaine variabilité de l'adaptation et de difficultés de croissance en MPL, les études (1) et (2) reposent sur des échantillons obtenus à partir de deux cultures indépendantes seulement. Pour l'étude (3), nous avons choisi du fait des difficultés rencontrées d'utiliser des cultures en MRS, un milieu complexe dans lequel la croissance de *L. bulgaricus* est plus reproductible.

2. Etudes en milieu chimiquement défini

2.1. Expression des régulateurs en condition d'adaptation à l'acide lactique

2.1.1. Introduction

L'analyse de la littérature montre que les analyses par 2DE apportent très peu, voire aucune information sur les régulateurs impliqués dans l'ATR. Cette observation peut s'expliquer de plusieurs façons : (i) les PM et pI des régulateurs pourraient être incompatibles avec leur détection par 2DE. L'analyse des régulateurs potentiels d'ATCC 11842 montre que les pI de 18 d'entre eux ne sont pas compatibles avec les études en 2DE. (ii) Les régulateurs sont souvent peu abondants et les CAI des 54 régulateurs de *L. bulgaricus*, compris entre 0,25 et 0,49 suggèrent qu'il en sera de même dans cette espèce. Seuls quatre régulateurs se trouvent parmi les gènes prédits comme étant surexprimés chez *L. bulgaricus*. (iii) Enfin, les régulateurs pourraient être inactivés ou activés en condition acide sans que leur expression ou leur abondance varient sensiblement.

Nous avons donc choisi d'étudier les modifications de l'expression des 55 régulateurs, 5 facteurs σ et les 5 systèmes à deux composants d'ATCC 11842 en condition d'adaptation au

stress acide par RT-PCR quantitative. Cette étude avait deux objectifs principaux : identifier les régulateurs impliqués dans l'ATR et leur régulons potentiels et identifier des promoteurs induits ou réprimés par l'acidité du milieu.

2.1.2. Mise au point des conditions expérimentales

Ce paragraphe est un complément au Matériel et Méthodes disponible p 165.

2.1.2.1. Conditions d'adaptation en MPL et préparation des échantillons

L'adaptation au stress acide est réalisée sur des cultures en phase exponentielle de croissance en les incubant à pH 5,25 pendant 90 min. Un test de survie à une épreuve acide (pH 4,2, 30 min) permet d'évaluer le niveau de tolérance avant et après adaptation (figure 16). Les échantillons d'ARN sont extraits à partir des culots cellulaires préparés en parallèle des tests physiologiques (figure 17). Les milieux à pH 5,25 et à pH 4 sont obtenus par acidification avec de l'acide lactique (voir matériel et méthodes p 165).

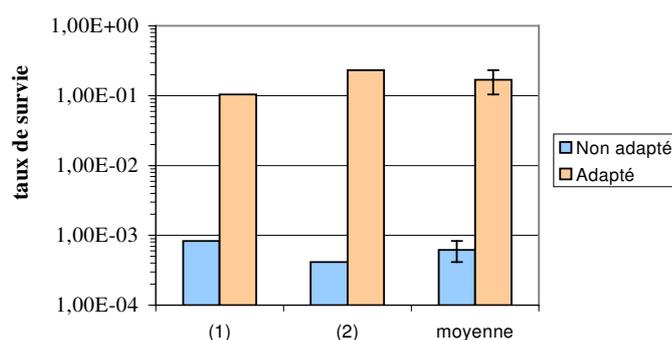


Figure 16 : test d'adaptation au stress acide

Taux de survie = (nb de colonies à T₃₀)/(nb de colonies à T₀)

Les ARN ont ensuite été transcrits en ADNc en utilisant des oligonucléotides aléatoires. Pour limiter les variations dues à l'efficacité des transcriptions inverses, quatre réactions ont été effectuées sur chaque échantillon avec 5 µg d'ARN avec la RT Powerscript™ de Clontech puis mélangées et aliquotées. Chaque aliquote n'a été utilisé que deux fois au maximum.

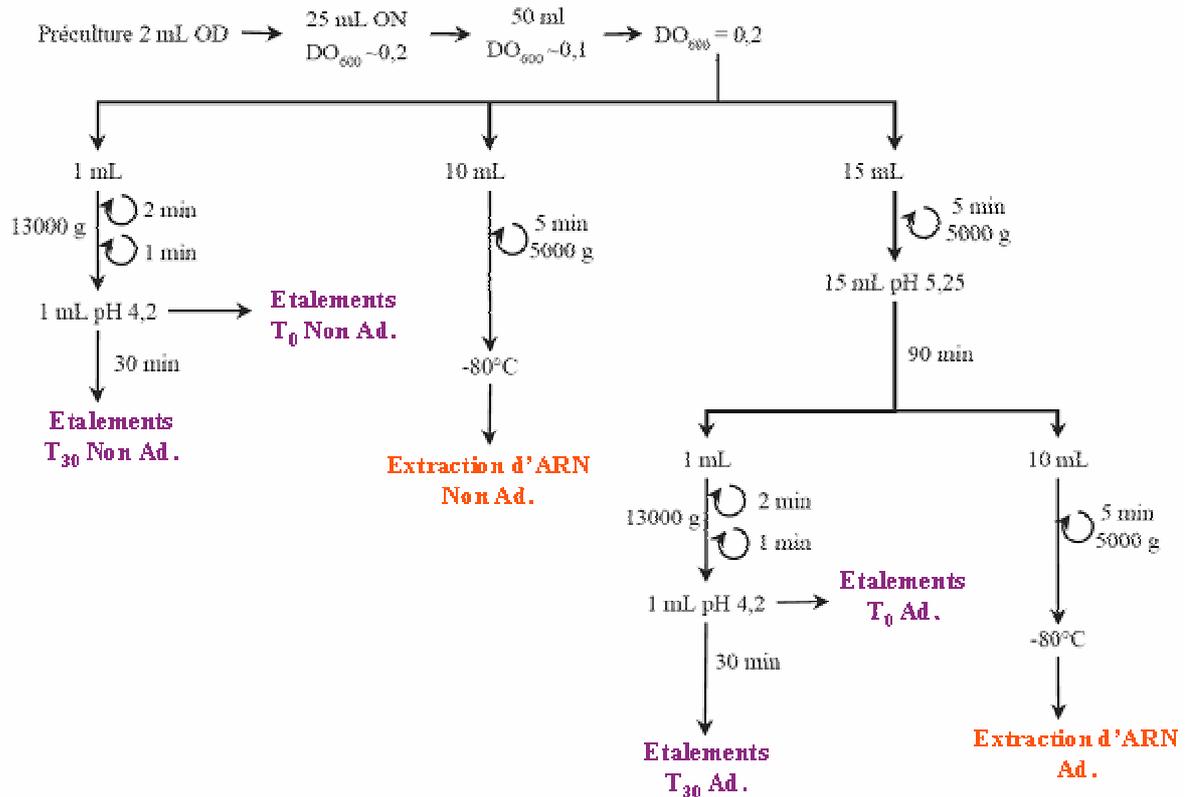


Figure 17 : Schéma d'expérience d'adaptation acide et d'obtention des échantillons pour la préparation d'ARN.

En violet, étalements permettant d'évaluer la survie des cellules à l'épreuve acide. En rouge, les culots cellulaires dont sont extraits les ARN utilisés dans l'étude.

OD : culture sur la journée ; ON : culture sur la nuit ; DO : Densité Optique ; Ad. : Adapté ; Non Ad. : Non Adapté

Pour plus de détails, voir matériel et méthodes p 165

2.1.2.2. PCR quantitative et gène de référence

Les mesures d'expression ont été réalisées sur les ARN issus des 2 expériences indépendantes et pour chaque gène considéré, les mesures en PCR quantitative sont réalisées en duplicata sur chacun des 4 échantillons (2 échantillons non adaptés et 2 échantillons adaptés). Les efficacités des couples d'oligonucléotides ont été déterminées par des dilutions en série d'ADN chromosomique. Nous avons sélectionné les couples pour lesquels l'efficacité est comprise entre 90 et 110 %. Lorsque ce travail a été effectué, nous ne disposions pas des courbes de fusion qui permettent de déterminer si l'amplification est bien spécifique, nous avons donc vérifié sur gel d'agarose que nous n'obtenions qu'un seul produit de taille compatible avec celle attendue.

La normalisation des quantités d'ADNc lors des PCR quantitatives a été effectuée en utilisant comme référence des oligonucléotides spécifiques du gène *hIbA*, qui code une protéine histone like, HU. D'après des expériences de Northern blot, l'expression de *hIbA* est constante au cours de la croissance d'ATCC 11842 (P. Serror, communication personnelle). L'efficacité de ce couple d'oligonucléotides sur ADN chromosomique est de 97 %. Lors des expériences de PCR quantitatives (q-PCR) permettant de déterminer la variation d'expression des régulateurs, les résultats obtenus pour *hIbA* montraient une variabilité importante entre les duplicata réalisés. Nous avons alors mesuré l'efficacité du couple d'oligonucléotides utilisé pour *hIbA* sur des dilutions en série d'ADNc. Dans ces conditions, l'efficacité du couple d'oligonucléotides n'est que de 60 %, ce qui implique qu'ils ne sont pas assez efficaces pour servir de référence. Cependant, les échantillons d'ADNc utilisés dans cette étude ont été utilisés dans une seconde étude avec un gène de référence valide (voir § 2.2 p 159). Les résultats de cette étude montrent qu'il n'y a pas de différence entre les quantités d'ADNc présentes dans les échantillons adaptés et non adaptés.

Je n'ai pas pu, dans le temps imparti, retester l'efficacité des oligonucléotides et la variation d'expression des 64 gènes testés en parallèle avec une référence valide. J'ai néanmoins choisi d'exposer ce travail en présentant les résultats de q-PCR sans les rapporter à la pseudo-référence interne.

2.1.3. Résultats

Les 64 gènes étudiés (49 régulateurs transcriptionnels, 5 facteurs σ et 5 histidine kinases et régulateurs de réponse des systèmes à deux composants) sont exprimés à des niveaux mesurables en RT-PCR quantitative dans les conditions adaptée et non adaptée. Nous n'avons observé aucune variation importante de l'expression des gènes entre ces deux conditions, le maximum de variation étant de 3,7 fois. Nous avons retenu toutes les variations d'un facteur supérieur à 2 dans au moins une des deux expériences indépendantes (tableau 12).

L'expression de 13 systèmes de régulation varie lors de l'adaptation à l'acide lactique, 4 seraient induits et 9 réprimés. Par analyse du contexte génomique, nous avons recherché les gènes qui pourraient être contrôlés par ces régulateurs. J'ai basé l'analyse sur (i) la proximité génétique et notamment les gènes transcrits en opéron et (ii) le type de fonctions habituellement contrôlées par le régulateur considéré. Dans 8 cas, l'expression des gènes potentiellement régulés a été mesurée avant et après adaptation acide (tableau 12). Les résultats ayant évolué selon la prise en compte ou non de la référence, les autres cibles potentielles ont été identifiées plus tard et n'ont pas été étudiées.

Tableau 12 : Régulateurs induits ou réprimés en condition d'adaptation à l'acide lactique et les gènes potentiellement régulés.

Famille	n°ldb	Rapport A/NA		Gène potentiellement contrôlé par le régulateur	n°ldb	Rapport A/NA	
		(1)	(2)			(1)	(2)
Crp/Fnr	0482	3,77	2,45	CPx ATPase (fragment)*	0478	2,23	2,66
TetR	2141	2,66	1,80	ABC transporter (putative antimicrobial peptides efflux)*	2142	2,67	1,78
TetR	2071	2,26	1,50	ABC transporter (putative antimicrobial peptides efflux)*	2070	2,07	3,61
MarR	0658	1,59	2,11	CPx ATPase*	0660	1,56	3,31
GntR	1689	0,40	0,53	ABC transporter (ATPase subunit fragment)*	1688-6	nd	nd
GntR	1715	0,40	0,42	Putative ABC transporter*	1714	nd	nd
MerR	1075°	0,51	0,34	Multidrug efflux protein	1074	nd	nd
ArgR	0810	0,42	0,40	Phosphopantothénocysteine synthetase/decarboxylase*	0811	nd	nd
Xre	0159	0,42	0,39	Putative ion transporter	0162	0,35	0,40
σ^{ECF}	1881	0,45	0,35	putative anti-sigma factor*	1880	0,45	0,49
Prt	0347	0,43	0,35	Purine metabolism		nd	nd
DeoR	2082	0,32	0,34	Fumarate hydratase	2086	0,36	0,23
DeoR	1182	0,33	0,21	Glycine/bétaine ABC transporter	1181	1,25	1,08

Sont indiqués les rapports d'expression entre la condition adaptée et non adaptée (A/NA) pour chaque expérience indépendante ((1) et (2)). Les régulateurs dont l'expression diminue en condition d'adaptation sont grisés.

*: gènes en opéron avec le régulateur ; nd : variation d'expression non déterminée.

° : et *ldb1132* (gènes identiques, voire dans le texte pour plus d'explications)

Le régulateur présentant le facteur d'induction le plus important (3,1 en moyenne) est un régulateur de la famille Crp/Fnr (*ldb0482*). Les régulateurs de cette famille sont impliqués dans le contrôle de fonctions variées dont les mécanismes de réponse aux stress environnementaux (pour revue : Korner *et al.*, 2003). La protéine Ldb0482 est homologue à FlpA de *L. lactis* et à Flp de *L. casei*, des régulateurs de la sous-famille Flp (pour Fnr like protein). L'activité de ces deux régulateurs est modulée par le potentiel redox et FlpA de *L. lactis* est impliqué dans la résistance au stress oxydant (Scott *et al.*, 2000). Chez *L. lactis*, FlpA contrôle l'import d'ions Zn^{2+} . Chez ATCC 11842, un gène codant une ATPase de transport d'ions métalliques se situe en opéron avec *ldb0482* (figure 18) mais il est inactivé par un changement de cadre de lecture (*ldb0477/0478*). En amont de cet opéron, nous avons détecté un motif TTGAT-N₄-ATCAA identique au site de liaison du régulateur Fnr d'*E. coli* et des deux régulateurs FlpA et FlpB de *L. lactis* (Korner *et al.*, 2003), ce qui pourrait indiquer que Ldb0482 est bien responsable de la régulation de cet opéron. La recherche de ce motif dans le génome nous a permis d'identifier une autre cible potentielle, le gène *msrB*, codant la protéin-méthionine-S-oxide réductase, impliquée dans la résistance au stress oxydant. Il serait intéressant de vérifier si ce gène est induit en condition d'adaptation à l'acidité.

Deux régulateurs de la famille TetR sont induits d'un facteur ~ 2 en condition d'ATR (*ldb2141* et *ldb2071*). Les régulateurs de la famille TetR sont généralement des répresseurs impliqués dans le contrôle de gènes associés à diverses fonctions dont la synthèse d'antibiotiques, la réponse au stress osmotique ou l'efflux de composés toxiques (pour revue : Ramos *et al.*, 2005). Les répresseurs de la famille TetR sont souvent transcrits de manière divergente par rapport à leurs cibles. Chez ATCC 11842, aucun des deux régulateurs ne présente de gènes adjacents orientés de cette façon (figure 18). Cependant, chaque régulateur est en opéron avec un ABC transporteur potentiellement impliqué dans l'efflux de peptides antimicrobiens, ce qui correspond aux fonctions classiquement régulées par cette famille de régulateur. Les régulateurs de la famille TetR ayant en général des régulons assez restreint, il est probable que ces régulateurs ne contrôlent que l'opéron auxquels ils appartiennent.

Le dernier régulateur à être induit en condition d'ATR (facteur inférieur à 2) appartient à la famille MarR (*ldb0658*). Cette famille de régulateur est impliquée dans le contrôle de fonctions de résistance aux composés toxiques ou de virulence (pour revue : Wilkinson et Grove, 2006). *Ldb0658* est homologue à CopY d'*E. hirae*, régulateur de l'homéostasie du cuivre chez cette bactérie (Solioz et Stoyanov, 2003). Chez ATCC 11842, *ldb0658* est en opéron avec un gène codant une ATPase de transport de métaux (*ldb0610*), homologue à CopA d'*E. hirae* (figure 18). Ce système étant largement décrit au paragraphe 2.2 de cette partie, je ne développerai pas sa description ici.

Parmi les gènes réprimés se trouvent deux gènes codant des régulateurs de la famille GntR, *ldb1689* et *ldb1715* réprimés d'un facteur 2,2 et 2,4 respectivement. Ces deux régulateurs appartiennent à la sous-famille YtrA des régulateurs de type GntR (Rigali *et al.*, 2002). Les régulateurs de cette sous-famille sont peu étudiés, le seul à avoir été caractérisé est YtrA de *B. subtilis*, régulateur d'un système de transport impliqué dans l'utilisation d'acétoïne (Yoshida *et al.*, 2000). Les régulateurs de type YtrA sont en général codés en opéron avec un ABC transporteur (Rigali *et al.*, 2002). Le gène *ldb1715* est transcrit dans un opéron comprenant quatre gènes : *ldb1715*, deux gènes codant un ABC transporteur (*ldb1714* et *ldb1713*) et un gène codant une perméase d'acides aminés (*aapA*) (figure 18). En amont de *ldb1715*, un motif GTGTACTA_{n4}-AATAgTACAgT, similaire au site de liaison proposé pour YtrA de *B. subtilis* (différences en minuscules ; Yoshida *et al.*, 2000) et comprenant les résidus conservés entre les sites de liaison des différents membres de la famille YtrA (en gras ; Rigali *et al.*, 2002) est détecté, ce qui indiquerait que cet opéron est bien régulé par *Ldb1715*.

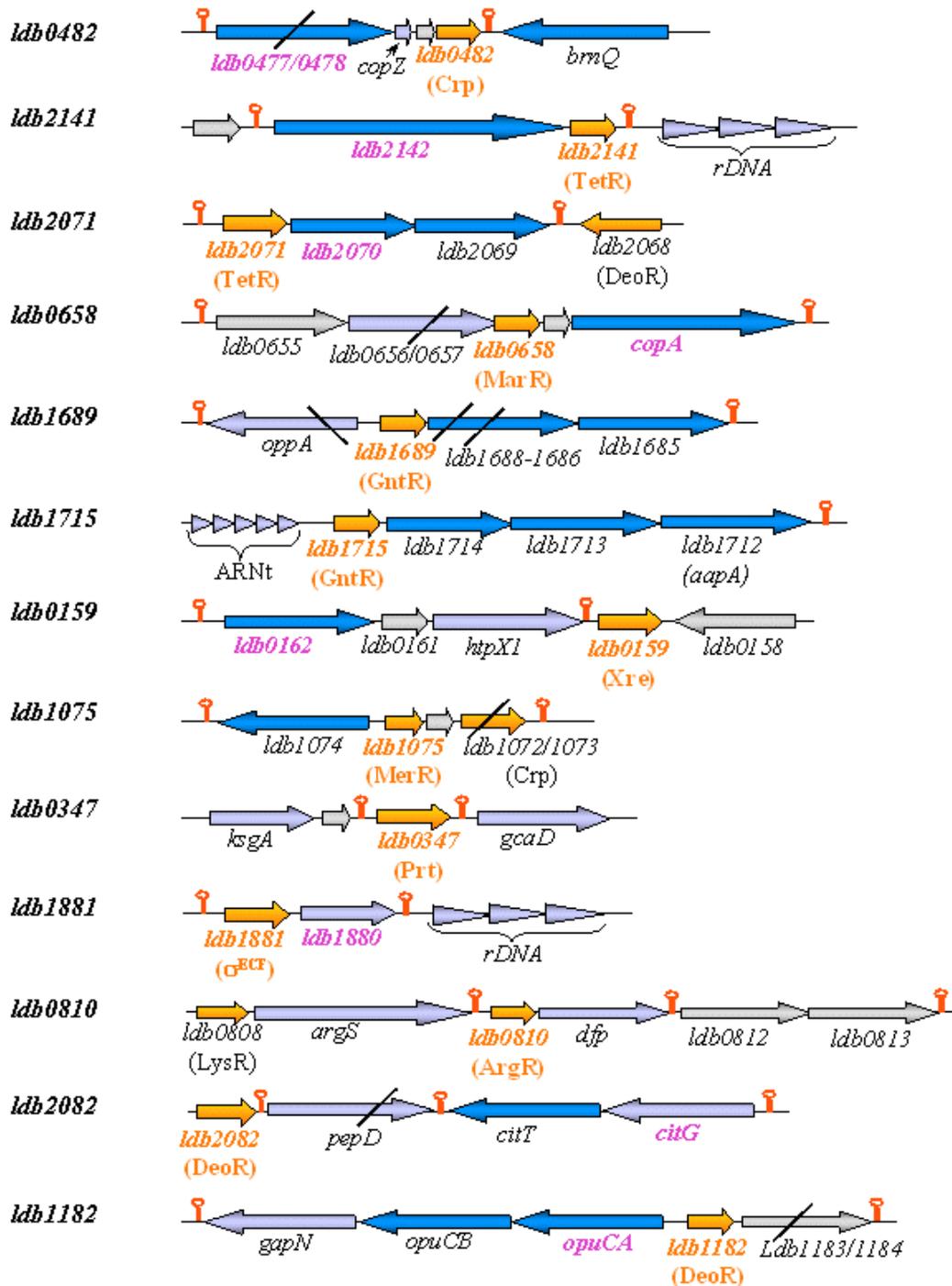


Figure 18 : Contexte génétique des régulateurs dont l'expression est affectée lors de l'ATR.

Les régulateurs dont l'expression est affectée sont indiqués en orange, ainsi que la famille à laquelle ils appartiennent (entre parenthèses). Les gènes que nous avons définis comme cible potentielle et dont la variation d'expression a été mesurée lors de l'ATR sont indiqués en gras et en violet. Les gènes dont les produits sont impliqués dans des fonctions de régulation sont colorés en orange, dans les fonctions de transport en bleu, dans d'autres fonctions en mauve (voir tableau 12 pour fonctions) et ceux pour lesquels la fonction est inconnue, en gris. Les gènes barrés représentent des pseudogènes, le nombre de barres correspondant au nombre de mutations produisant une CDS supplémentaire. Les structures tige-boucle représentent les terminateurs de transcription.

° copie identique à *ldb1132* ainsi que contexte génétique identique.

Le gène *ldb1689* est lui aussi probablement transcrit en opéron avec deux gènes codant les sous-unités d'un ABC transporteur, mais le gène codant la sous-unité ATPasique est fragmenté (*ldb1688-1686* ; figure 18). En amont de *ldb1689*, le motif **TGTAaTAGAtAAatagtACA**, conservé en amont des orthologues chez *L. plantarum*, *L. johnsonii* et *L. acidophilus* et comprenant aussi les résidus du consensus défini par Rigali et coll. (en gras), est détecté. Ce motif pourrait être le site de liaison du régulateur Ldb1689.

ATCC 11842 contient deux gènes codant des régulateurs de type MerR. Ces gènes se situent dans la répétition inversée du génome et sont donc absolument identiques, ainsi que leur contexte génétique. Durant l'ATR, les gènes *ldb1075* et *ldb1132* (codant les deux copies du régulateur) sont réprimés d'un facteur 2,3. Les régulateurs de la famille MerR sont des activateurs qui contrôlent des fonctions liées aux réponses à des stimuli environnementaux comme le stress oxydant, les métaux lourds ou les antibiotiques (pour revue : Brown *et al.*, 2003). Dans certains cas, le régulateur est transcrit par un promoteur divergent et chevauchant avec celui de sa cible. Chez ATCC 11842, *ldb1075* et *ldb1132* sont transcrits de manière divergente avec un gène codant une protéine d'efflux de composés toxiques (*ldb1074/1133* ; figure 18). La séquence intergénique fait 92 pb de long, ce qui indique que les promoteurs du régulateur et de la protéine d'efflux sont probablement chevauchants. La variation d'expression du gène codant la protéine d'efflux n'a pas été mesurée, il est difficile de prévoir si ce gène est induit ou réprimé. En effet, chez les régulateurs de type MerR étudiés, la régulation sur le promoteur du gène cible a été étudiée mais pas la régulation du régulateur lui-même.

Le gène *ldb0810* est réprimé d'un facteur 2,4. Il code un régulateur de type ArgR, impliqué dans la régulation de la biosynthèse ou de la dégradation de l'arginine (Makarova *et al.*, 2001). Chez ATCC 11842, les seuls gènes conservés avec ceux qui sont régulés par ArgR chez *L. plantarum* (Nicoloff *et al.*, 2004) sont *argH* (mais sous forme de pseudogène) et des homologues de *carAA* et *carAB*. Le gène *ldb0810* est transcrit en opéron avec un gène impliqué dans le métabolisme du panthoténate (*dfp*, *ldb0811* ; figure 18), dont l'expression n'a pas été mesurée en condition d'ATR.

Le gène *ldb0159* est réprimé d'un facteur 2,5. Il code pour un régulateur de type Xre, dont les membres sont en général des régulateurs transcriptionnels de bactériophages. Il existe peu de données sur le type de fonctions régulées. Le gène *ldb0159* ne se situe pas dans un prophage et semble transcrit seul. L'unité transcriptionnelle située en amont est composée de trois gènes codant une perméase d'ions putative, une protéine hypothétique et une protéase liée à la paroi, HtpX (Sakoh *et al.*, 2005), (figure 18). Nous avons mesuré la variation d'expression pour le

premier gène de cet opéron, il est, comme le régulateur, réprimé d'un facteur 2,7. Cependant, nous ne pouvons pas affirmer que Ldb0159 est le régulateur de l'opéron situé en amont.

Un gène codant un facteur σ^{ECF} (*ldb1881*) est réprimé d'un facteur 2,5. Les facteurs σ^{ECF} sont impliqués dans la régulation d'une grande variété de fonctions en réponse à des stimuli environnementaux (pour revue : Helmann, 2002). Les facteurs σ^{ECF} sont généralement transcrits en opéron avec des gènes codant des protéines régulatrices de son activité. Le gène *ldb1881* est transcrit en opéron avec un anti-facteur σ , *ldb1880*, réprimé lui aussi d'un facteur ~2. Par cette analyse, nous ne pouvons pas déterminer quelles sont les cibles potentielles de ce σ^{ECF} .

Le gène *ldb0347*, réprimé d'un facteur 2,6, code pour une protéine de la famille PRT comprenant des enzymes de la voie de biosynthèse des nucléotides et des régulateurs transcriptionnels de cette même voie (pour revue : Sinha et Smith, 2001). Ldb0347 est homologue à PurR, régulateur des gènes de biosynthèse des purines chez *B. subtilis* (Weng *et al.*, 1995) et *L. lactis* (Kilstrup et Martinussen, 1998). Chez ces deux bactéries, le mécanisme d'action diffère : PurR est un répresseur des gènes de synthèse des purines chez *B. subtilis* (Weng *et al.*, 1995) alors que chez *L. lactis*, il agit comme activateur de ces mêmes gènes mais comme auto-répresseur (Kilstrup et Martinussen, 1998). ATCC 11842 possède l'opéron *pur* (*purEKBCSQLFMNHD*), identique à celui de *B. subtilis*, ainsi que le gène *purA*, isolé dans le génome. Ces deux unités transcriptionnelles sont, chez *B. subtilis*, sous contrôle de PurR. Nous n'avons pas testé la variation d'expression de ces gènes, ce qui ne nous permet pas de prédire s'ils sont induits ou réprimés en condition d'ATR, du fait du mode d'action différent de PurR chez les deux bactéries à Gram+ chez lequel il a été étudié.

Enfin, deux gènes, *ldb2082* et *ldb1182*, codant pour des régulateurs appartenant à la famille de régulateur de DeoR sont réprimés d'un facteur 3 et 3,7 respectivement. Les régulateurs de type DeoR sont des répresseurs impliqués dans le contrôle du métabolisme carboné (Elgrably-Weiss *et al.*, 2006). Le produit du gène *ldb2082* est homologue à CitI de *L. lactis* et *W. paramesenteroides*, impliqué dans la régulation du métabolisme du citrate (Martin *et al.*, 2005). Contrairement à ce qui est observé chez *L. bulgaricus* où *citI* est réprimé, l'expression de *citI* est induite à pH 5 chez *L. lactis* biovar *diacetylactis* (Martin *et al.*, 2004). Contrairement aux autres membres de la famille DeoR, CitI a été montré comme étant un activateur de la transcription de ces gènes cibles, impliqués dans l'utilisation du citrate et notamment ceux codant la citrate lyase (Martin *et al.*, 2005). *L. bulgaricus* ne possède pas les gènes du métabolisme du citrate. A proximité de *ldb2082*, dans une unité transcriptionnelle indépendante, deux gènes *citG* (codant une fumarate hydratase) et *citF* (codant un antiport 2-oxoglutarate/malate), non homologues aux

gènes *cit* régulés par *citI*, sont détectés (figure 18). La mesure de l'expression de *citG* montre que ce gène est réprimé d'un facteur 3,4, équivalent à celui de *ldb2082*.

L'autre régulateur de la famille DeoR dont l'expression est réprimée, Ldb1182 ne présente pas d'homologue étudié chez d'autres bactéries. En général, les régulateurs DeoR ont des régulons restreints et sont liés génétiquement à leur cible (Barriere *et al.*, 2005 ; Campos *et al.*, 2004 ; van Rooijen et de Vos, 1990). Le gène en aval de *ldb1182* code pour une protéine de fonction inconnue et est inactivé (figure 18). En amont de *ldb1182* se trouve un opéron contenant trois gènes, deux codant le système de transport de glycine-bétaine OpuC et le dernier codant GapN, enzyme intervenant dans le métabolisme central carboné. L'expression de *opuCA* ne varie pas en condition d'ATR, suggérant que cet opéron ne fait pas partie du régulon de Ldb1182.

Parmi les huit cibles potentielles dont l'expression a été mesurée, sept ont une expression qui varie de la même façon que celle du régulateur considéré. Cela n'est guère surprenant pour six d'entre elles qui font a priori partie du même opéron que le régulateur. Pour ces six cas, ce résultat constitue cependant une confirmation des résultats obtenus pour le régulateur.

Il s'est avéré que la technique de PCR quantitative n'était pas très adaptée pour suivre l'expression d'un aussi grand nombre de gènes. Considérant que nos résultats pointaient plusieurs systèmes de transport, nous avons choisi de cibler notre étude sur les transporteurs d'ions détectés dans le génome. En effet, les transports de molécules chargées à travers la membrane contribuent au potentiel de membrane ($\Delta\Psi$), une composante, avec le ΔpH , de la force protomotrice. Un autre intérêt de cette étude était de revalider, avec un gène de référence, certaines des mesures d'expression obtenues dans cette étude.

2.2. Etude de l'expression des systèmes de transport d'ions

2.2.1. Détermination du gène de référence

Pour cette étude, nous avons utilisé comme gène de référence le gène *ldb1245* qui code pour la deuxième copie du gène *sigA*. Lors de l'étude sur les régulateurs, l'expression de *ldb1245* ne semblait pas affectée par l'adaptation acide. Nous avons donc décidé de vérifier qu'il pouvait servir de gène de référence.

L'efficacité du couple d'oligonucléotides déterminée sur ADN chromosomique est de 102%. Nous avons déterminé l'efficacité de ce couple sur les ADNc utilisés au cours de l'étude. Sur

ADNc, il est de 105% sur une gamme de dilutions de $1/180$, $1/540$, $1/1620$ et $1/4860$. La dilution d'ADNc utilisée pour la détermination de la variation d'expression est de $1/1000$, et est donc comprise dans la gamme de linéarité. De plus, les courbes de fusion montrent qu'un seul produit est détecté pour l'amplification par ce couple d'oligonucléotides.

Pour rappel de la technique, j'exposerai ici la méthode de calcul des facteurs d'induction en tenant compte d'une référence (Livak et Schmittgen, 2001). Le cycle auquel la fluorescence dépasse la valeur seuil détermine la valeur de cycle seuil (Ct). La mesure de la variation de l'expression d'un gène lors de l'adaptation acide fait intervenir les équations suivantes :

- (i) $\Delta Ct = Ct$ du gène en condition adaptée (Ad)- Ct du gène en condition non adaptée (NA)
- (ii) $\Delta\Delta Ct = \Delta Ct$ du gène d'intérêt - ΔCt du gène de référence
- (iii) Rapport d'expression Ad/NA = $2^{-\Delta\Delta Ct}$

Pour déterminer si l'expression de *ldb1245* est affectée lors de l'adaptation, nous avons mesuré la variation d'expression du gène *ldb1245* (ΔCt) pour des échantillons issus d'une même expérience physiologique. Pour cela, chaque échantillon a été déposé 3 fois et le Ct pris en compte est la moyenne des Ct obtenus pour les trois dépôts identiques. Pour éliminer des variations dues aux dilutions des ADNc effectuées, l'expérience a été réalisée trois fois et nous avons fait la moyenne des ΔCt obtenus pour chaque expérience physiologique. Pour l'expérience (1), le ΔCt moyen obtenu est de +0,03, et pour l'expérience (2), il est de -0,02. Ces résultats nous ont permis de déterminer que l'expression de *ldb1245* n'est pas affectée par l'adaptation à l'acidité et qu'il pouvait donc servir de gène de référence.

2.2.2. Expression des transporteurs d'ions lors de l'ATR

Cette partie fait l'objet d'un manuscrit soumis à Applied and Environmental Microbiology.

Résumé

L'adaptation aux conditions acides est un aspect important pour la physiologie de *Lactobacillus bulgaricus*, une des deux bactéries impliquées dans la production de yaourt. Pour survivre sous ces conditions, les bactéries lactiques utilisent divers mécanismes qui ont pour but de limiter l'acidification du cytoplasme et de prévenir ou réparer les dommages causés aux composants cellulaires.

Nous avons analysé la séquence génomique de *L. bulgaricus* et nous avons pu mettre en évidence que peu de fonctions impliquées dans le contrôle du pH et décrites chez les autres bactéries lactiques sont présentes, à part l' H^+ ATPase. Les systèmes de transport d'ions ont été

impliqués dans l'adaptation à l'acidité dans quelques bactéries n'appartenant pas aux bactéries lactiques. Dans cet article, nous présentons une étude de l'expression des systèmes de transport d'ions de *L. bulgaricus* durant l'adaptation acide. De manière frappante, les systèmes les plus induits sont trois ATPases présentant le motif CPx des transporteurs de métaux lourds. L'homologie de séquence et la présence d'un motif commun dans les régions promotrices des gènes codant ces trois ATPases suggèrent fortement qu'elles soient impliquées dans l'homéostasie du cuivre. L'induction de ce système pourrait aider la cellule à éviter les dommages indirects qui pourraient résulter de l'acidification du milieu..

Induction of heavy metal transporting CPx type ATPases during acid adaptation in *Lactobacillus bulgaricus*

S. Penaud, A. Fernandez, S. Boudebbouze, S.D. Ehrlich, E. Maguin*, and M. van de Guchte

INRA, Génétique Microbienne, 78352 Jouy en Josas Cedex, France

*corresponding author

E. Maguin

Génétique Microbienne

INRA - CRJ

78352 Jouy en Josas cedex

France

tel: +33 1 34 65 25 18

fax: +33 1 34 65 25 21

e-mail: emmanuelle.maguin@jouy.inra.fr

Running title: Acid adaptation in *Lactobacillus bulgaricus*

Section: Food Microbiology

Abstract

Adaptation to acidic conditions is an important aspect of the physiology of *Lactobacillus bulgaricus*, one of the key bacteria in yogurt fermentation. In order to survive under these conditions, lactic acid bacteria (LAB) can employ a variety of mechanisms that aim to limit the acidification of the cytoplasm and to prevent or repair damage to cellular components. We analyzed the genome sequence of *L. bulgaricus* and found few of the pH control functions that have been described in other LAB, however, apart from an H⁺ ATPase. Ion transport systems have been implicated in acid adaptation in some non LAB bacteria. Here we present a study of the expression of ion transport systems during acid adaptation in *L. bulgaricus*. Intriguingly, the most significantly induced were three ATPases carrying the CPx signature of heavy metal transporters. Protein homology and the presence of a conserved sequence motif in the promoter regions of the genes encoding these proteins strongly suggest that they are involved in copper homeostasis. Induction of this system is thought to assist in avoiding indirect damage that could result from medium acidification.

Introduction

Lactobacillus delbrueckii ssp. *bulgaricus* (*L. bulgaricus*) belongs to the lactic acid bacteria (LAB), a taxonomically heterogeneous group of bacteria characterized by the production of lactic acid and concomitant acidification of their environment. Within this group, *L. bulgaricus* is one of the economically most important representatives owing to its worldwide use in the production of yogurt, the product of milk fermentation by *L. bulgaricus* and *Streptococcus thermophilus*.

Traditionally, yogurt consumption has been credited with health-beneficial effects (2), and some of these have been shown to depend on the presence of live bacteria (45, 53). In addition, depending on the country national laws can require yogurt to contain at least 10 million live bacteria per gram at the time of marketing. This makes bacterial survival during fermentation, product handling and shelf life an important issue. An important property of yogurt that affects bacterial survival is its acidity. During yogurt fermentation, lactic acid production causes the pH to fall to about 4.2, down from an initial value of 6.6 to 6.8 in milk (FAO).

Functions that have been implicated in acid adaptation in LAB include those involved in changes in the cell envelope and repair of DNA and protein damage that often play a role in the response to other stress factors as well (for recent reviews see (13, 19, 52)). In addition, mechanisms that play an active role in the control of cellular pH and proton motive force have been described. These generally include H⁺ expelling proteins or protein complexes, notably the H⁺ ATPase, and enzymes involved in the production of alkaline compounds.

A less studied group of proteins that may play a role in acid adaptation consists of ion transport systems. A chloride channel (23) and a K⁺ uptake system (15) have been implicated in acid resistance in *E. coli* and *S. mutans*, respectively, and attributed a role in the equilibration of electric charges when the H⁺ ATPase or other systems are proceeding to a net efflux of protons.

Here we address the question of how *L. bulgaricus* adapts to the low pH it creates in the culture medium during growth, and escapes the fate of many spoilage bacteria that are inhibited or killed in this hostile environment. We present an inventory of acid adaptation mechanisms known from other LAB in the genome of *L. bulgaricus*, and a study of the expression of ion transport systems during acid adaptation.

Materials and methods

Bacterial strain and growth conditions. The bacterial strain used throughout this study is *L. delbrueckii* ssp. *bulgaricus* ATCC11842, obtained from the American Type Culture

Collection. *L. bulgaricus* was cultivated at 42°C in chemically defined medium (MPL, (12)) containing 5% lactose and 50mM MES (pH 6.5). Additional vitamins (cyanocobalamine 1mg/L, orotic acid 5mg/L, 2-deoxythymidine 5mg/L, DL-6,8-thioctic acid 2.5mg/L and pyridoxamine 5 mg/L) were added to this medium where indicated (MPLm), to improve growth. Colony forming units (cfu) were counted after plating on MRS agar medium (Difco Laboratories, Inc., Detroit, MI) and incubation for 42 hours at 42°C under anaerobic conditions.

Determination of acid tolerance and adaptation. A single fresh colony was used to inoculate 1 mL of MPL and incubated for 8 hours at 42°C without shaking. The culture was then diluted 1000-fold in 13 ml of prewarmed MPL medium, and incubation was continued overnight. The resulting culture ($OD_{600} < 0.2$) was diluted to an OD_{600} of 0.05 with prewarmed MPL and incubated until the OD_{600} reached 0.2. To measure acid tolerance, cells from 1 ml were harvested by centrifugation at $13000 \times g$ (2 min at room temperature), and the pellet was resuspended in 1 ml of MPL adjusted to pH 4.0 or 4.2 with lactic acid. Immediately after resuspension and after 30 min at 42°C, aliquots of serial dilutions in MRS were spread on MRS agar medium. Survival was calculated by dividing the number of colony forming units (CFU) per ml after incubation at pH 4.0 or 4.2 by the number of CFU per ml immediately after resuspension.

For experimental acid adaptation, the culture with an OD_{600} of 0.2 was divided in two aliquots. For one aliquot (non adapted sample), acid tolerance was measured as described above. Cells from 1 ml of the second aliquot were harvested at $13000 \times g$ and resuspended in 1 ml of MPL adjusted to pH 5.25 with lactic acid, and incubated at 42°C for 90 min (acid adapted sample). Then, acid tolerance was measured as described above. The acid adaptation factor was calculated as the survival of acid adapted cells divided by the survival of non adapted cells.

In order to monitor the natural acid adaptation during growth, a single fresh colony was used to inoculate 1 mL of MPL and incubated for 8 hours at 42°C without shaking. The culture was then diluted 1000-fold in 25 ml of prewarmed MPL medium, and incubation was continued overnight. The resulting culture ($OD_{600} < 0.2$) was diluted to an OD_{600} of 0.05 with prewarmed MPL. At 1 hour intervals, OD_{600} and pH were measured and acid tolerance was determined as above.

A slightly modified protocol was used for acid adaptation experiments described in Table 4, where cultures were grown in MPLm, subjected or not to a sublethal pH of 4.75 in MPLm acidified with either HCl or lactic acid (adaptation), and tested for acid tolerance in MPLm acidified to a pH of 3.6 with HCl.

RNA extraction and cDNA preparation. For extraction of RNA, cells from 10 ml of culture with an OD_{600} of ~ 0.2 , with or without acid adaptation treatment, were harvested at 2600

× g (10 min at 4°C). The pellet was resuspended in 100 µl of cold 10 mM Tris, 1 mM EDTA buffer (pH 7.5), frozen in liquid nitrogen, and stored at –80°C. The cells were thawed on ice and broken by shaking with 0.6 g of glass beads (<106 µm in diameter, Sigma) (3 times 45 s) in a Bio 101 Fastprep apparatus (Savant Instruments, Holbrook, NY) in the presence of 500 µl acid phenol. 1 ml of Trizol (Life technologies, Rockville, MD) was added to the aqueous phase and used to extract RNA as recommended by the supplier. DNA was removed using the DNA-free system (Ambion, Austin, TX) (0.04 u Dnase 1 /µg ARN, 30 min at 37°C), and the absence of DNA was checked by PCR. RNA was quantified by measurement of the OD₂₆₀. cDNA was obtained by reverse transcription of 5µg of DNA-free RNA using Powerscript Reverse Transcriptase (Clontech Laboratories, Inc., Mountain View, CA) with 250 ng of random hexamers (New England Biolabs, Ipswich, MA) in a volume of 20 µl, as recommended by the supplier.

Quantitative PCR. Primers for Quantitative PCR (Table 1) were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). PCR was carried out in a 25 µl volume containing 10 µl of 1000-fold diluted cDNA, specific primers (0.2 µM each), and 12.5 µl ABsolute QPCR SYBR Green mix (ABgene, Epsom, UK). Thermo cycling was performed using an ABI 7700 instrument (Applied Biosystems, Foster city, CA) using the following parameters: 1 cycle at 95°C for 15 min, and 40 cycles at 94°C for 15 s and at 60°C for 1 min. Melting curve amplification was performed with 1,5 °C temperature increments per minute in order to check that the amplification did not produce secondary products. PCR efficiency was checked for each primer set using serial dilutions of cDNA. Cycle threshold (Ct) values were defined as the cycle number at which the fluorescence exceeded a fixed threshold value above the baseline. Induction of gene expression during acid adaptation, relative to gene expression before adaptation, was calculated using the comparative $\Delta\Delta C_t$ method (30) with *ldb1245* as the internal control gene for normalization of the amount of cDNA in the reaction. *ldb1245* encodes a sigma A homologue and was chosen after verification that its expression level was not significantly affected by acid adaptation in the two independent physiological experiments: for a given amount of total RNA, the amount of *ldb1245* mRNA varied by a factor inferior to 1.4 between acid adapted and non adapted cells (data not shown). All quantitative PCR measurements were realized in triplicate on samples derived from two independent cultures.

Transcription start mapping and motif search. Transcription starts were mapped using the 5'/3' RACE kit (Roche Applied Science, Mannheim, Germany) as recommended by the supplier. Mapping was realized with RNA extracted from acid adapted and non adapted cells, using the primers listed in Table 1. Conserved sequence motifs were searched within the 500 bp

upstream of the translation initiation codon using iMoMi (N. Pons *et al.* in preparation), based on the MEME algorithm (7).

Results

Evolution of acid tolerance during growth. During growth, LAB produce lactic acid and thereby gradually acidify the culture medium. In the course of this process, the bacteria adapt to the changing environment and become more tolerant to low pH. For *L. bulgaricus*, the extent of this adaptation during growth in a chemically defined medium (MPL) was evaluated by measuring the survival to an acid challenge (pH 4.0 for 30 min) (Fig.1). During the first seven hours of growth, the pH of the culture gradually decreased from 6.4 to 5.4, and the survival after acid challenge fluctuated between 0.01 % and 0.1 %. Then, while the pH of the culture further decreased to 5.1 after 8 hours and 4.8 after 9 hours, the survival to acid challenge rapidly increased to 4% and 78%, respectively. These results clearly illustrate that during growth and acidification of the culture medium, *L. bulgaricus* acquires an improved tolerance to acid challenge. The onset of a rapid development of acid tolerance coincided with a medium pH between 5.4 and 5.1, and a culture OD₆₀₀ between 0.65 and 0.85.

Experimental acid adaptation of *L. bulgaricus*. To dissociate the effect of adaptation to an acidic environment from potential growth phase related effects, we examined whether the exposure of an exponentially growing culture with an OD₆₀₀ of 0.2 to a low but sublethal pH (5.25) could induce an improved survival to acid challenge (30 min pH 4.0). The results presented in Fig. 2 show that such a response could indeed be induced, and that the adaptation factor increased with the time of incubation at pH 5.25. After 90 min, the survival to acid challenge was 1000-fold higher than in the non adapted control sample.

Inventory of genes involved in acid tolerance and ion transporters. An analysis of the genome sequence of *L. bulgaricus* ATCC11842 (M. van de Guchte *et al.*, submitted for publication) reveals only few of the functions that have been implicated in acid tolerance in other LAB (Table 2). Of the enzymes that have been described to produce alkaline compounds and thereby play a role in pH control, the genome only encodes two ornithine decarboxylases. *L. bulgaricus* does not possess an arginine deiminase pathway, no glutamate, histidine or aspartate decarboxylase, and no urease. Metabolic pathways for malolactic fermentation or citrolactic fermentation are not present either.

In contrast, of the H⁺ and other ion transporters that have been described to play a role in acid adaptation, *L. bulgaricus* encodes the H⁺-ATPase, a chloride channel, and a K⁺ uptake permease (Table 3). An inventory of ion transporters encoded in the genome revealed 16

additional transporters of which it may be envisaged that some play a role in acid adaptation.

Induction of ion transporters during acid adaptation. In order to identify ion transporters involved in acid adaptation, *L. bulgaricus* was grown in MPL until an OD₆₀₀ of 0.2, and part of the culture was acid adapted using conditions based on the results described above (90 min at pH 5.25). The development of acid tolerance in response to this treatment was confirmed by largely improved survival after acid challenge (data not shown). RNA was extracted from adapted and non adapted samples, and the relative expression of transporters was determined by reverse transcription followed by quantitative PCR (RT-QPCR).

Transcripts were detected for all genes tested, in both adapted and non adapted cells. The results presented in Table 3 show that among the 19 transporters tested, one (Ldb1239) was induced 31 to 45 fold during adaptation, and three others were induced by a factor superior to 2 in two repetitions of the experiment with independent cultures. Remarkably, these include the three CPx-type ATPases (Ldb1239, Ldb0660 and Ldb2158) present in the *L. bulgaricus* genome. CPx -type ATPases constitute a subclass of the P-type ATPases that is specialized in the transport of heavy metal ions (5, 44, 49). The fourth gene is a cation:proton antiporter (*ldb0226*).

The expression of six of the studied genes was repressed by a factor of about two. These include a branched chain amino acid:proton symporter (Ldb0483) and a Na⁺/H⁺antiporter (Ldb1810). Also repressed were a permease of unknown specificity (Ldb0162), an ammonia transporter (Ldb1663), a phosphate ABC transporter (Ldb0956-Ldb0959), and a cation transporting P-type ATPase (Ldb0456). Neither the H⁺-ATPase (Ldb0705-Ldb0712) nor the K⁺ uptake permease (Ldb0219) were induced or repressed. Induction of the chloride channel (Ldb2201) was low (1.7 fold), but reproducible.

Transcription start site mapping and motif search. To determine whether a common regulator might control the different transporter genes, the 5' ends of transcripts from the genes of which the expression varied by a factor 2 or more were determined. For genes that make part of an operon, as judged on the basis of the genome sequence, the transcription start site of the operon was determined. Experiments were performed using RNA isolated from adapted and non-adapted culture samples, and transcription start sites proved to be identical under both conditions indicating that the modulation of transcript levels did not rely on switching between two promoters. Start sites could be determined for all genes but one (*ldb0956*), and near consensus sigma A promoter sequences were identified upstream of all these sites (Fig. 3). *Ldb0660* appeared to be transcribed from two promoters, one directly upstream of the *ldb0658-0659-0660* operon and one situated further upstream, preceding *ldb0655* (Fig. 4A). The gene directly upstream of the *ldb0658-0659-0660* operon is a pseudogene (*ldb0656/0657*) and it seems likely that in the course of the degradation of this gene the accompanying transcription

terminator has also been affected (Fig. 4A), resulting in readthrough of transcription. *Ldb0655-0656/0657* transcription was not induced during acid adaptation (data not shown), but as a consequence of readthrough transcription from this operon, the apparent induction factor measured for the *ldb0658-0659-0660* operon will underestimate the induction of transcription initiation from the *ldb0658* promoter.

The promoter regions were subsequently analyzed with iMOMi (N. Pons *et al.*, in preparation) to identify shared motifs that could be indicative for coordinated regulation by a common regulator. A common motif, ykACArwyrTAArC comprising an internal inverted repeat (Fig. 4BC), could only be found, in one or two copies, in the promoter regions upstream of the CPx ATPase genes *ldb1239*, *ldb0660* and *ldb2158*. This motif was not found elsewhere in the *L. bulgaricus* genome. Nearly identical motifs were found in the promoter regions of homologues of *ldb1239* and *ldb0660*, but not *ldb2158*, in a number of closely and more distantly related other LAB (Fig 4D).

Characterization of CPx type ATPases. While the implication of several of the induced or repressed transporters in pH control is very well conceivable, we focussed on the CPx-type ATPases for further analysis because 1) all three intact ATPases of this type in the genome were induced, 2) one of these was very strongly induced, 3) a common sequence motif could be identified in the promoter regions of these genes, and 4) as described below, at least two of these proteins probably make part of a coordinated system.

Ldb1239 and *Ldb0660* are homologous to the copper transporting CopB and CopA proteins of *Enterococcus hirae* (36), respectively. The alignment of these proteins clearly illustrates the conservation of the TGE, DKTGT and GDGINDAP motifs common to P-type ATPases, and the CPx motif of the subgroup of heavy metal transporting CPx type ATPases (49) (Fig. 5). The conserved ALGLA membrane channel motif directly following the CPx motif strongly suggests that *Ldb1239* and *Ldb0660* are involved in the transport of monovalent Cu or Ag ions (49, 51). Of these, Cu seems the most relevant as this metal is present in low concentrations in milk (42). The CPxALGLA motif is only partially conserved in *Ldb2158*. Overall protein homology and a shared sequence motif in the promoter region, which may indicate a common regulation mechanism, suggest that this protein too may be involved in copper transport, however, although transport of another heavy metal ion can not be excluded.

Ldb0660 and *Ldb2158* lack the N-terminal domain that is present in most other heavy metal ATPases (Fig. 5). This difference is not considered to be important, however, as a similar lack has been observed in homologous proteins from several other bacteria, and in *E. coli* ZntA this domain was shown not to be essential for function or specificity (34).

Role of CPx type ATPases during acid adaptation. These observations raise the question

of the significance of the induction of CPx type ATPases, and more precisely a putative copper homeostasis control system, during acid adaptation. As copper is essentially known for its function as a cofactor in redox enzymes, one may wonder whether the slightly different history of acid adapted and non-adapted cells in terms of aeration may be responsible for the induction of the transporters. For acid adaptation, cells were submitted to centrifugation and resuspended in fresh medium (with a lower pH), while non adapted cells were frozen immediately after centrifugation (cf Materials and Methods). Therefore, a control experiment was performed in which non-adapted cells were resuspended in fresh medium (without pH change) and incubated for 90 min at 42 °C. This treatment did not induce the expression of the CPx type ATPases (data not shown), thereby eliminating this explanation.

A second explanation could be that lactate, used to acidify the adaptation medium, binds copper ions (49) thereby changing the effective copper concentration and inducing the system. Therefore, acid adaptation experiments were repeated using a medium of which the pH had been adjusted with HCl instead of lactic acid. The results shown in Table 4 demonstrate that expression of the CPx type ATPases was also induced under these conditions, thereby eliminating the possibility of a lactate effect and emphasizing the role of the pH of the medium.

Finally, adaptation experiments were also performed using a medium without added copper. While MPL(m) contains 0.01 μ M of CuSO₄, in this experiment *L. bulgaricus* was cultured, and adaptation was performed, in MPLm where this medium component had been omitted. The results presented in Table 4 show that this omission did not affect the induction of the CPx ATPases during acid adaptation.

Discussion

Acid adaptation is a complex process, which can mobilize a large spectrum of different cellular functions. In the current picture of this process, these comprise functions involved in the control of intracellular pH and proton motive force that would prevent the deleterious effects of acidification, and functions that repair damage to proteins and DNA once occurred.

A survey of the *L. bulgaricus* genome sequence learned that this bacterium possesses few of the proteins that have been implicated in the first class of functions in other bacteria. We therefore focussed on the potential role of proton and ion transport functions, other than the H⁺-ATPase that has previously been studied in detail in other LAB (15, 22, 25, 37) and is also present in *L. bulgaricus*. Transporters likely involved in acid adaptation were revealed by measuring their induction or repression in response to a pH shift. Although modification of gene expression under these conditions is not a prerequisite for a transporter to be involved in

adaptation, this approach allowed the identification of a number of transporters with a potential role in this process. These include several proton symporters and antiporters, and an ammonia transporter.

Surprisingly, the results of this study showed that acidification of the culture medium also, and notably, induced the enhanced expression of putative heavy metal transporting CPx type ATPases. Ldb0660 and Ldb1239 appeared homologous to CopA and CopB from *E. hirae*, responsible for the import and export, respectively, of Cu⁺ (48, 56). Copper is at the time important as a cofactor in redox enzymes and very toxic, necessitating a tight regulation of its intracellular concentration (49). In *E. hirae*, the *copA* and *copB* genes make part of one operon which also encodes a chaperone (CopZ) and a repressor (CopY), and is believed to ensure copper homeostasis. CopY represses the expression of the operon by binding to the operator sequence TACAnnTGTA (55). If copper is in excess, it will bind to CopY, thereby prevent the binding of CopY to the operator, and induce expression of the operon (39, 50). A shortage of copper also induces the expression of the operon, but the mechanism of this induction is not understood (36, 55). Paradoxically, either condition thus induces both the importer CopA and the exporter CopB, a feature which has been hypothesized to act as a safety mechanism to avoid copper intoxication in case of sudden variations in ambient copper availability (49).

The CopA and CopB homologues in *L. bulgaricus* share essential sequence motifs with the proteins from *E. hirae* and make part of the same phylogenetic group of transporters, two features that strongly suggest that the *L. bulgaricus* proteins too are involved in copper transport (44, 49, 51). The presence of a CopY homologue (Ldb0658) in *L. bulgaricus*, encoded by the gene upstream of and in operon with the *copA* homologue *ldb0660*, and the strong resemblance between the conserved motif upstream of the CPx type ATPase genes in *L. bulgaricus* and the *E. hirae* CopY binding site, reinforce this hypothesis. Further support comes from the presence of a *copY* pseudogene upstream of the *CopB* homologue *ldb1239*. A protein resembling CopZ (Ldb0480) is encoded elsewhere in the genome (data not shown).

In *L. bulgaricus*, *ldb0660* and *ldb1239* are situated in different transcription units which would theoretically allow a fine-tuning between import and export functions. On the other hand, the CopA and Ldb0660 orthologue from *Escherichia coli* has been described as an exporter rather than an importer (17, 43), and we can therefore not exclude the possibility that in *L. bulgaricus* both Ldb1239 and Ldb0660 are involved in copper export.

The results presented in this study strongly suggest that acidification of the culture medium affects copper homeostasis in *L. bulgaricus* and thereby threatens cell viability. Induction of the three CPx ATPase genes was observed after acidification with HCl as well as with lactic acid. The mechanism underlying this induction is as yet not clear, however. The strong resemblance of

the ATPases, the CopY homologue and the putative operator sites of the three separate transcription units with the respective elements of the intensively studied *E. hirae* system suggest that CopY plays a key role in this induction. It is not clear if and how copper could play a role in the induction, as this also takes place in a culture medium that only contains trace amounts of copper if any at all. It is tempting to believe that CopY operator fixation is also sensitive to other stimuli than copper concentration, directly related to medium acidity.

Transport activity of the *E. hirae* CopB protein is strongly pH dependent, and its efficiency rapidly decreases below pH 6 (48). If this property is shared by the *L. bulgaricus* Cop homologues, an increased expression at low pH values could serve to compensate reduced efficiency and ensure copper homeostasis. This system could thus be involved in avoiding indirect damage as a consequence of medium acidification that could have dramatic effects for the cell if not properly addressed.

Several recent studies in other bacteria evaluated the changes in gene expression during growth or short time exposure to acidic conditions using a genome wide transcriptomic approach (*Lactobacillus plantarum* (38); *Helicobacter pylori* (4, 33, 54); *Mycobacterium tuberculosis* (18); *Streptococcus pneumoniae* (32)). Interestingly, one of these studies identified CopA as strongly induced in the highly acid resistant stomach colonizing bacterium *Helicobacter pylori*, after 48 hr of growth at pH 5.5 (compared to a control grown at pH 7.2) (4).

Transcriptomic and proteomic (e.g. (29)) analyses of acid adaptation rapidly reveal the induction or repression of a multitude of new genes, in addition to those classically implicated. Part of these changes are expected or can be understood intuitively (e.g. modifications of metabolic activity), while others are more surprising, but not necessarily less important. They allow us to further develop our understanding of the profound and pleiotropic effect that acidification of the culture medium, brought about by LAB, has on the physiology of the bacterial cell.

Acknowledgments

We thank I. Guillouard and C. Grimaldi for useful discussions. We thank N. Galleron, B. Quinquis, V. Brachet, M.-C. Beaussart, J. Musset, and L. Prioux for technical assistance. The work of S.P. was in part financed by Danone Vitapole

References

1. **Abe, K., H. Hayashi, and P. C. Maloney.** 1996. Exchange of aspartate and alanine. Mechanism for development of a proton-motive force in bacteria. *J Biol Chem* **271**:3079-84.
2. **Adolfsson, O., S. N. Meydani, and R. M. Russell.** 2004. Yogurt and gut function. *Am J Clin Nutr* **80**:245-56.
3. **Altermann, E., W. M. Russell, M. A. Azcarate-Peril, R. Barrangou, B. L. Buck, O. McAuliffe, N. Souther, A. Dobson, T. Duong, M. Callanan, S. Lick, A. Hamrick, R. Cano, and T. R. Klaenhammer.** 2005. Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc Natl Acad Sci U S A* **102**:3906-12.
4. **Ang, S., C. Z. Lee, K. Peck, M. Sindici, U. Matrubutham, M. A. Gleeson, and J. T. Wang.** 2001. Acid-induced gene expression in *Helicobacter pylori*: study in genomic scale by microarray. *Infect Immun* **69**:1679-86.
5. **Axelsen, K. B., and M. G. Palmgren.** 1998. Evolution of substrate specificities in the P-type ATPase superfamily. *J Mol Evol* **46**:84-101.
6. **Azcarate-Peril, M. A., E. Altermann, R. L. Hoover-Fitzula, R. J. Cano, and T. R. Klaenhammer.** 2004. Identification and inactivation of genetic loci involved with *Lactobacillus acidophilus* acid tolerance. *Appl Environ Microbiol* **70**:5315-22.
7. **Bailey, T. L., and C. Elkan.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**:28-36.
8. **Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols.** 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* **22**:1554-8.
9. **Bolotin, A., P. Wincker, S. Mauger, O. Jaillon, K. Malarme, J. Weissenbach, S. D. Ehrlich, and A. Sorokin.** 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* **11**:731-53.
10. **Boyd, D. A., D. G. Cvitkovitch, A. S. Bleiweis, M. Y. Kiriukhin, D. V. Debabov, F. C. Neuhaus, and I. R. Hamilton.** 2000. Defects in D-alanyl-lipoteichoic acid synthesis in *Streptococcus mutans* results in acid sensitivity. *J Bacteriol* **182**:6055-65.
11. **Chen, Y. Y., C. A. Weaver, and R. A. Burne.** 2000. Dual functions of *Streptococcus salivarius* urease. *J Bacteriol* **182**:4667-9.
12. **Chervaux, C., S. D. Ehrlich, and E. Maguin.** 2000. Physiological study of *Lactobacillus delbrueckii* subsp. *bulgaricus* strains in a novel chemically defined medium. *Appl Environ Microbiol* **66**:5306-11.
13. **Cotter, P. D., and C. Hill.** 2003. Surviving the acid test: responses of gram-positive bacteria to low pH. *Microbiol Mol Biol Rev* **67**:429-53, table of contents.
14. **Cox, D. J., and T. Henick-Kling.** 1989. Chemiosmotic energy from malolactic fermentation. *J Bacteriol* **171**:5750-2.
15. **Dashper, S. G., and E. C. Reynolds.** 1992. pH regulation by *Streptococcus mutans*. *J Dent Res* **71**:1159-65.
16. **Diaz-Torres, M. L., and R. R. B. Russell.** 2001. HtrA protease and processing of extracellular proteins of *Streptococcus mutans*. *Fems Microbiology Letters* **204**:23-28.
17. **Fan, B., G. Grass, C. Rensing, and B. P. Rosen.** 2001. *Escherichia coli* CopA N-terminal Cys(X)(2)Cys motifs are not required for copper resistance or transport. *Biochem Biophys Res Commun* **286**:414-8.
18. **Fisher, M. A., B. B. Plikaytis, and T. M. Shinnick.** 2002. Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *J Bacteriol* **184**:4025-32.

19. **Foster, J. W.** 2004. Escherichia coli acid resistance: tales of an amateur acidophile. *Nat Rev Microbiol* **2**:898-907.
20. **Fozo, E. M., R. G. Quivey, Jr., R. Faustoferri, K. Monahan, and R. Marquis.** 2004. The fabM gene product of Streptococcus mutans is responsible for the synthesis of monounsaturated fatty acids and is necessary for survival at low pH. *J Bacteriol* **186**:4152-8.
21. **Hanna, M. N., R. J. Ferguson, Y. H. Li, and D. G. Cvitkovitch.** 2001. uvrA is an acid-inducible gene involved in the adaptive response to low pH in Streptococcus mutans. *Journal of Bacteriology* **183**:5964-5973.
22. **Hutkins, R. W., and N. L. Nannen.** 1993. pH Homeostasis in Lactic Acid Bacteria. *J. Dairy Sci.* **76**:2354-2365.
23. **Iyer, R., T. M. Iverson, A. Accardi, and C. Miller.** 2002. A biological role for prokaryotic CIC chloride channels. *Nature* **419**:715-8.
24. **Kleerebezem, M., J. Boekhorst, R. van Kranenburg, D. Molenaar, O. P. Kuipers, R. Leer, R. Tarchini, S. A. Peters, H. M. Sandbrink, M. W. Fiers, W. Stiekema, R. M. Lankhorst, P. A. Bron, S. M. Hoffer, M. N. Groot, R. Kerkhoven, M. de Vries, B. Ursing, W. M. de Vos, and R. J. Siezen.** 2003. Complete genome sequence of Lactobacillus plantarum WCFS1. *Proc Natl Acad Sci U S A* **100**:1990-5.
25. **Kobayashi, H.** 1985. A proton-translocating ATPase regulates pH of the bacterial cytoplasm. *J Biol Chem* **260**:72-6.
26. **Kobayashi, H., T. Suzuki, and T. Unemoto.** 1986. Streptococcal cytoplasmic pH is regulated by changes in amount and activity of a proton-translocating ATPase. *J Biol Chem* **261**:627-30.
27. **Koch, H. G., M. Moser, and M. Muller.** 2003. Signal recognition particle-dependent protein targeting, universal to all kingdoms of life. *Rev Physiol Biochem Pharmacol* **146**:55-94.
28. **Lemos, J. A. C., and R. A. Burne.** 2002. Regulation and physiological significance of ClpC and ClpP in Streptococcus mutans. *Journal of Bacteriology* **184**:6357-6366.
29. **Lim, E. M., S. D. Ehrlich, and E. Maguin.** 2000. Identification of stress-inducible proteins in Lactobacillus delbrueckii subsp. bulgaricus. *Electrophoresis* **21**:2557-61.
30. **Livak, K. J., and T. D. Schmittgen.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**:402-8.
31. **Magni, C., D. de Mendoza, W. N. Konings, and J. S. Lolkema.** 1999. Mechanism of citrate metabolism in Lactococcus lactis: resistance against lactate toxicity at low pH. *J Bacteriol* **181**:1451-7.
32. **Martin-Galiano, A. J., K. Overweg, M. J. Ferrandiz, M. Reuter, J. M. Wells, and A. G. de la Campa.** 2005. Transcriptional analysis of the acid tolerance response in Streptococcus pneumoniae. *Microbiology* **151**:3935-46.
33. **Merrell, D. S., M. L. Goodrich, G. Otto, L. S. Tompkins, and S. Falkow.** 2003. pH-regulated gene expression of the gastric pathogen Helicobacter pylori. *Infect Immun* **71**:3529-39.
34. **Mitra, B., and R. Sharma.** 2001. The cysteine-rich amino-terminal domain of ZntA, a Pb(II)/Zn(II)/Cd(II)-translocating ATPase from Escherichia coli, is not essential for its function. *Biochemistry* **40**:7694-9.
35. **Molenaar, D., J. S. Bosscher, B. ten Brink, A. J. Driessen, and W. N. Konings.** 1993. Generation of a proton motive force by histidine decarboxylation and electrogenic histidine/histamine antiport in Lactobacillus buchneri. *J Bacteriol* **175**:2864-70.
36. **Odermatt, A., H. Suter, R. Krapf, and M. Solioz.** 1993. Primary structure of two P-type ATPases involved in copper homeostasis in Enterococcus hirae. *J Biol Chem* **268**:12775-9.
37. **O'Sullivan, E., and S. Condon.** 1999. Relationship between acid tolerance, cytoplasmic pH, and ATP and H⁺-ATPase levels in chemostat cultures of Lactococcus lactis. *Appl Environ Microbiol* **65**:2287-93.
38. **Pieterse, B., R. J. Leer, F. H. Schuren, and M. J. van der Werf.** 2005. Unravelling the multiple effects of lactic acid stress on Lactobacillus plantarum by transcription profiling. *Microbiology* **151**:3881-94.

39. **Portmann, R., D. Magnani, J. V. Stoyanov, A. Schmechel, G. Multhaupt, and M. Solioz.** 2004. Interaction kinetics of the copper-responsive CopY repressor with the cop promoter of *Enterococcus hirae*. *J Biol Inorg Chem* **9**:396-402.
40. **Pridmore, R. D., B. Berger, F. Desiere, D. Vilanova, C. Barretto, A. C. Pittet, M. C. Zwahlen, M. Rouvet, E. Altermann, R. Barrangou, B. Mollet, A. Mercenier, T. Klaenhammer, F. Arigoni, and M. A. Schell.** 2004. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc Natl Acad Sci U S A* **101**:2512-7.
41. **Quivey, R. G., R. C. Faustoferri, K. A. Clancy, and R. E. Marquis.** 1995. Acid Adaptation in *Streptococcus-Mutans* Ua159 Alleviates Sensitization to Environmental-Stress Due to ReCa Deficiency. *Fems Microbiology Letters* **126**:257-261.
42. **Renner, E.** 1989. *Micronutrients in milk and milk-based food products.* Elsevier Applied Science, London.
43. **Rensing, C., B. Fan, R. Sharma, B. Mitra, and B. P. Rosen.** 2000. CopA: An *Escherichia coli* Cu(I)-translocating P-type ATPase. *Proc Natl Acad Sci U S A* **97**:652-6.
44. **Rensing, C., M. Ghosh, and B. P. Rosen.** 1999. Families of soft-metal-ion-transporting ATPases. *J Bacteriol* **181**:5891-7.
45. **Rizkalla, S. W., J. Luo, M. Kabir, A. Chevalier, N. Pacher, and G. Slama.** 2000. Chronic consumption of fresh but not heated yogurt improves breath-hydrogen status and short-chain fatty acid profiles: a controlled study in healthy men with or without lactose maldigestion. *Am J Clin Nutr* **72**:1474-9.
46. **Rollan, G., G. L. Lorca, and G. Font de Valdez.** 2003. Arginine catabolism and acid tolerance response in *Lactobacillus reuteri* isolated from sourdough. *Food Microbiology* **20**:313-319.
47. **Sanders, J. W., K. Leenhouts, J. Burghoorn, J. R. Brands, G. Venema, and J. Kok.** 1998. A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. *Mol Microbiol* **27**:299-310.
48. **Solioz, M., and A. Odermatt.** 1995. Copper and silver transport by CopB-ATPase in membrane vesicles of *Enterococcus hirae*. *J Biol Chem* **270**:9217-21.
49. **Solioz, M., and J. V. Stoyanov.** 2003. Copper homeostasis in *Enterococcus hirae*. *FEMS Microbiol Rev* **27**:183-95.
50. **Strausak, D., and M. Solioz.** 1997. CopY is a copper-inducible repressor of the *Enterococcus hirae* copper ATPases. *J Biol Chem* **272**:8932-6.
51. **Tottey, S., P. R. Rich, S. A. Rondet, and N. J. Robinson.** 2001. Two Menkes-type atpases supply copper for photosynthesis in *Synechocystis* PCC 6803. *J Biol Chem* **276**:19999-20004.
52. **van de Guchte, M., P. Serron, C. Chervaux, T. Smokvina, S. D. Ehrlich, and E. Maguin.** 2002. Stress responses in lactic acid bacteria. *Antonie Van Leeuwenhoek* **82**:187-216.
53. **Van de Water, J., C. L. Keen, and M. E. Gershwin.** 1999. The influence of chronic yogurt consumption on immunity. *J Nutr* **129**:1492S-5S.
54. **Wen, Y., E. A. Marcus, U. Matrubutham, M. A. Gleeson, D. R. Scott, and G. Sachs.** 2003. Acid-adaptive genes of *Helicobacter pylori*. *Infect Immun* **71**:5921-39.
55. **Wunderli-Ye, H., and M. Solioz.** 1999. Effects of promoter mutations on the in vivo regulation of the cop operon of *Enterococcus hirae* by copper(I) and copper(II). *Biochem Biophys Res Commun* **259**:443-9.
56. **Wunderli-Ye, H., and M. Solioz.** 2001. Purification and functional analysis of the copper ATPase CopA of *Enterococcus hirae*. *Biochem Biophys Res Commun* **280**:713-9.
57. **Yamashita, Y., T. Takehara, and H. K. Kuramitsu.** 1993. Molecular characterization of a *Streptococcus mutans* mutant altered in environmental stress responses. *J Bacteriol* **175**:6220-8.

Table 1. Primers used in this study

Gene	Quantitative PCR primers		Transcriptional start mapping primers
<i>ldb0162</i>	F*	AAGTGGACCCTGCCTTACCT	R GCAGAAGATGACCGTTGTGA
	R*	GCAGAAGATGACCGTTGTGA	R TGGTCAAGCAGGTAATGTCGT
			R AGACTCGATCGTGTGCCAGT
<i>ldb0219</i>	F	TGCTGCCTCTCCCTGAACT	
	R	TGGCATAGAAGGGGTTGATT	
<i>ldb0226</i>	F	ACCTGCACGACTCCTGGTAT	R AAGTATCCAGCTTCCGGTCA
	R	AAGTATCCAGCTTCCGGTCA	R GCTTGCCGCTCTCTGACT
			R CCATCAAGATCCCGACGA
<i>ldb0341</i>	F	CACCGGGATCAAGAACATCT	
	R	TCAAGGCATCAGTCCAGTTG	
<i>ldb0426</i>	F	ATGGCCGGTGGAAAAGTCTA	
	R	ATCATGGCAAAACGGATGAA	
<i>ldb0456</i>	F	GACCTGCGTCTTTTGGTAGC	R TGCTGGAGTAAGCCATGTTG
	R	TGCTGGAGTAAGCCATGTTG	R CGCGTTCAAAAAGCACAACCTA
			R AAGCGGGCAAACATTGAAC
<i>ldb0483</i>	F	CCGGCCTCTTATTTTCTCA	R AGAAACTTGCCGATCCAGTC
	R	AGAAACTTGCCGATCCAGTC	R CTGGGCAAAAGAGCTGAGAT
			R AGTAGCTGTAAGTGTGGCTGACC
<i>ldb0655</i>			R AGGCCAGGTCATAGAATTGC
			R GTAGCCCTTGATACCCGACA
			R GGTCCGTTCAAAGCAGTCATAG
<i>ldb0657</i>	F	GCCGGTCTCTTTTCCACTCT	
	R	GGTCCGTTCAAAGCAGTCATAG	
<i>ldb0658</i>	F	CAGGCAAATAGCAGCGAGAT	R CCATTCACCTTCCATATGCTC
	R	ATACGAAGCGGTGAACCTTG	R GGAGCCGCTCTTCGCCTTT
			R CAGCAGGGTCTTGACAGTTG
<i>ldb0660</i>	F	GATGATCTGGGAATCATGC	
	R	GGCGATTAGGCTGTTTATGT	
<i>ldb0705</i>	F	CCTGCAGAACTATGCACGTC	
	R	GGACTTCACCCGAAAAGATA	
<i>ldb0956</i>	F	TCGTGTCAATCAACGGTGTTA	R CTTTGGCTGGCCTTTAGTGT
	R	CTTTGGCTGGCCTTTAGTGT	R TTGCCGGTGAAAACCTTTGAT
			R GTTGATGATCGGAGCGATG
<i>ldb1239</i>	F	GTATTCCAGCCGATGGTGT	R TGACCTTATCGCCAACCTTTCTT
	R	TGACCTTATCGCCAACCTTTCTT	R TTCGGTCATATCGTGATCCA
			R ATTCCTTCTCGCCAACAACC
<i>Ldb1245</i>	F	GTCTGCTCTGGCCCTCAAGG	
	R	TCCGAACCGAGTCGTAAACCT	
<i>ldb1301</i>	F	ATGATCGGAACCGACAAGAC	
	R	AGGCAAAACCGATGATTGTC	
<i>ldb1306</i>	F	GTGACGTTGTTCCATCGTG	R CCGATCAAGGAGACGATGAA
	R	AAGAAGAGGCCGTTGGTGTA	R AAGAAGAGGCCGTTGGTGTA
			R CACGATGGAAAACAACGTCAC
<i>ldb1663</i>	F	CCGGTATCGTTGGCTCTTTA	R TGCCTGCTTGCCCTAAGAGTT
	R	TGCCTGCTTGCCCTAAGAGTT	R CGAAGACCAATAGAAAAGTTGAACC
			R AGGCCACGAATAAGACGATG
<i>ldb1810</i>	F	GGACCTTGATTGCTGTCTGG	R AAGACCGAGGGAACGAAGAA
	R	AAGACCGAGGGAACGAAGAA	R ACAGCAATCAAGGTCCCAAT
			R CCAGAAGATGAGTAAGGCGATG
<i>ldb1895</i>	F	CGCCTGCAGACAGAGTTTTA	
	R	CCAGACTGGCACGTCATAGA	
<i>ldb2158</i>	F	GGTGGCCTTCACTGGTCTTA	R GGTAATGGCACGGCGTAG
	R	GTTTTGGATTCTTCGGTTCG	R GATCATCCATTACGCCAGTA
			R CCTTCTTTCAAGTCCCCAAAC
<i>ldb2159</i>	F	GTCCTCAGCCTCTTACCAC	
	R	AAGACGAAGTAGAGGGCTTCAA	
<i>ldb2201</i>	F	TTGATCTTGAGCTTGAACCAAA	
	R	CAGGCCAGAATCGTAAGACAC	

* F, Forward; R, Reverse.

Table 2. Genes involved in acid tolerance in LAB.

Gene	Function	gene number in <i>L. bulgaricus</i>	Reference
Protein degradation			
<i>clpP</i>	ATP-dependent Clp protease proteolytic subunit	<i>ldb0624</i>	(28)
<i>htrA</i>	HtrA-like serine protease	<i>ldb0140</i>	(16)
DNA repair			
<i>recA</i>	Recombinase A	<i>ldb0599</i>	(41)
<i>uvrA</i>	Excinuclease ABC, subunit A	<i>ldb0615</i>	(21)
Cell envelope			
<i>fabM</i>	Fatty acid desaturase	np	(20)
<i>dagK</i>	Diacylglycerol kinase	np	(57)
<i>ffh</i>	Signal recognition particle protein Ffh	<i>ldb1372</i>	(27)
<i>dltC</i>	D-alanyl carrier protein	<i>ldb2146</i>	(10)
Proton pumps and related functions			
<i>atpA-H</i>	H ⁺ transporting ATPase / ATP synthase	<i>ldb0705-ldb0712</i> *	(26)
<i>gadCD</i>	Glutamate decarboxylase	np	(47)
<i>adc</i>	Aspartate decarboxylase	np	(1)
<i>hdc</i>	Histidine decarboxylase	np	(35)
<i>odc</i>	Ornithine décarboxylase	<i>ldb0547, ldb1775</i> **	(6)
<i>mlf</i>	Malolactic fermentation	np	(14)
<i>citC-G, M, P</i>	Citrolactic fermentation	np	(31)
Ammonia production			
<i>arcABC</i>	Arginine deiminase	np	(46)
<i>ureA-G</i>	Urease	np	(11)

Only functions that have been studied in detail, beyond an observation of induction in proteomic or transcription studies, are presented. np, not present. *Different subunits of the H⁺ transporting ATPase are encoded by *ldb0705* to *ldb0712*. **The *L. bulgaricus* genome encodes two ornithine décarboxylases.

Table 3. Induction of ion transporters in *L. bulgaricus* ATCC11842.

Gene	Function	TC number	mRNA ratio*	
			(1)	(2)
<i>ldb1239</i>	CPx-type ATPase, putative Cu transporter CopB	3.A.3.5.2	31.1	44.7
<i>ldb0660</i>	CPx-type ATPase, putative Cu transporter CopA	3.A.3.5.1	2.1	3.5
<i>ldb2158</i>	CPx-type ATPase	3.A.3.6.5	2.6	2.1
<i>ldb0226</i>	Cation:proton antiporter	2.A.37.1.1	2.5	2.1
<i>ldb0341</i>	Cation transporting P-type ATPase	3.A.3.4.1	2.1	1.2
<i>ldb2201</i>	Chloride channel	1.A.11.5.1	1.8	1.6
<u><i>ldb0426-0424</i></u>	Putative Cobalt ABC transporter	3.A.1.23.2	1.0	1.1
<u><i>ldb0705-0712</i></u>	H ⁺ transporting ATPase	3.A.2.1.1	1.1	0.8
<i>ldb0219</i>	K ⁺ uptake permease	2.A.72.1.1	0.8	1.0
<i>ldb2159</i>	Na ⁺ /H ⁺ antiporter	2.A.37.2.1	0.8	0.7
<i>ldb1301</i>	Cation transporting P-type ATPase	3.A.3.3.3	0.7	0.6
<i>ldb1895</i>	Putative cation transporter	1.A.35.3.1	0.6	0.5
<i>ldb1306</i>	Cation efflux protein, probable metal/H ⁺ antiporter	2.A.4.1.1	0.5	0.6
<i>ldb0456</i>	Cation transporting P-type ATPase	3.A.3.2.4	0.5	0.5
<i>ldb1810</i>	Na ⁺ /H ⁺ antiporter	2.A.35.1.1	0.5	0.5
<u><i>ldb0956-0959</i></u>	Phosphate ABC transporter	3.A.1.7.1	0.6	0.4
<i>ldb0483</i>	Branched chain amino acid:H ⁺ symporter	2.A.26.1.3	0.5	0.5
<i>ldb1663</i>	Ammonia transporter	2.A.49.1.1	0.5	0.5
<i>ldb0162</i>	Putative ion transporter	2.A.47.5.1	0.4	0.3

For multi-component transporters encoded by an operon, the first and the last gene of the operon are indicated, and genes of which the expression was measured are underlined. TC number, classification number of the nearest homologue in the Transport Classification Database. *acid adapted cells / non adapted cells, measured by RT-QPCR; results for two independent experiments (1, 2) are presented, each figure represents the mean of three QPCR measurements.

Table 4. Effect of different adaptation media on induction of CPx type ATPases.

Gene	Lactic acid		HCl	
	+ Cu	- Cu	+ Cu	- Cu
<i>ldb1239</i>	32.7	34.6	20.5	21.2
<i>ldb0660</i>	14.5	9.0	3.3	2.3
<i>ldb2158</i>	5.7	4.5	4.2	3.9

Figures represent the induction factor (relative gene expression [acid adapted / non-adapted]), measured by RT-QPCR for the conditions indicated. Induction factors for lactic acid can not be compared to those in Table 3 because of differences in media composition (MPLm was used instead of MPL, cf Materials and Methods) and pH of adaptation. +Cu, MPLm (containing 0.01 μ M CuSO₄); -Cu, MPLm from which CuSO₄ was omitted. Absolute values of induction factors obtained with HCl and with lactic acid can not be compared because these conditions represent qualitatively and quantitatively different stress levels for the cell.

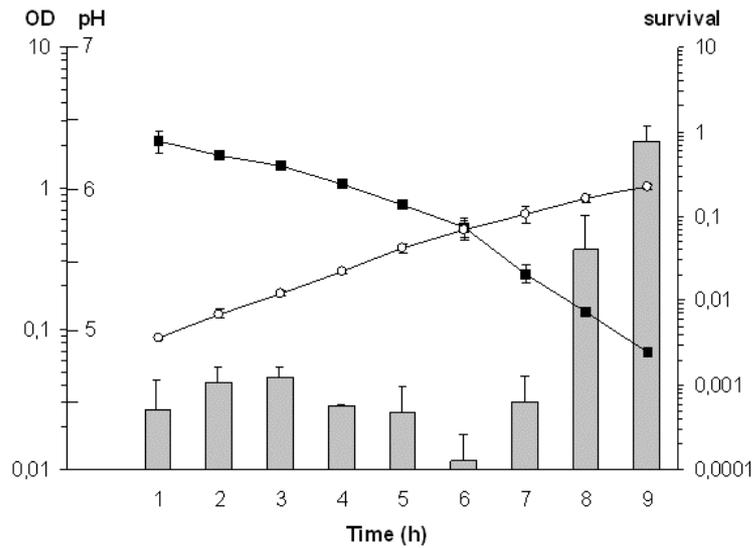


Fig. 1. Evolution of optical density at 600 nm (OD), pH and acid tolerance during growth in MPL.

Data represent the mean of three independent experiments. Standard deviations are indicated by error bars. ○, OD at 600 nm; ■, pH. Histogram, survival after acid challenge (pH 4.0).

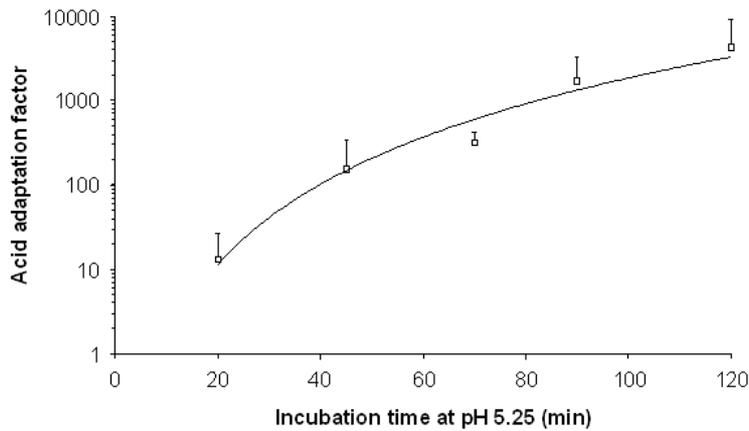


Fig. 2. Acid adaptation factor as a function of adaptation time.

Acid tolerance (survival to pH 4.0) of an exponentially growing culture was measured after different periods of incubation at pH 5.25, and compared to acid tolerance of the non adapted culture. Survival of the non adapted culture varied between 8×10^{-6} and 2.6×10^{-4} , with a mean of 8×10^{-5} . Data represent the mean for five independent experiments. Standard deviations are indicated by error bars.

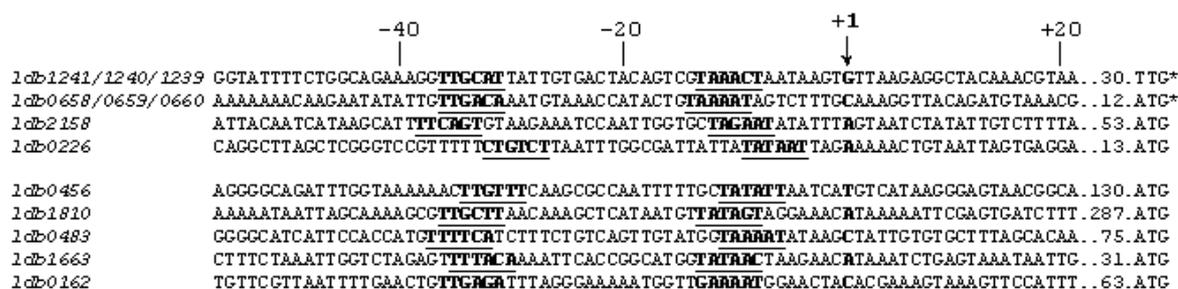


Fig. 3. Transcription start sites.

Transcription start sites (+1) and putative –35 and –10 promoter sequences (underlined) are presented for a selection of ion transport genes from Table 3. The number of nucleotides between the presented sequence and the translation start codon shown in the right part of the figure is indicated. Putative ribosome binding sites could be identified upstream of the indicated start codons (data not shown). *Genes *ldb1239* and *ldb0660* make part of operons (cf Fig. 4). In this case the transcription start site of the operon is presented, and the distance to the start codon of the first gene in the operon.

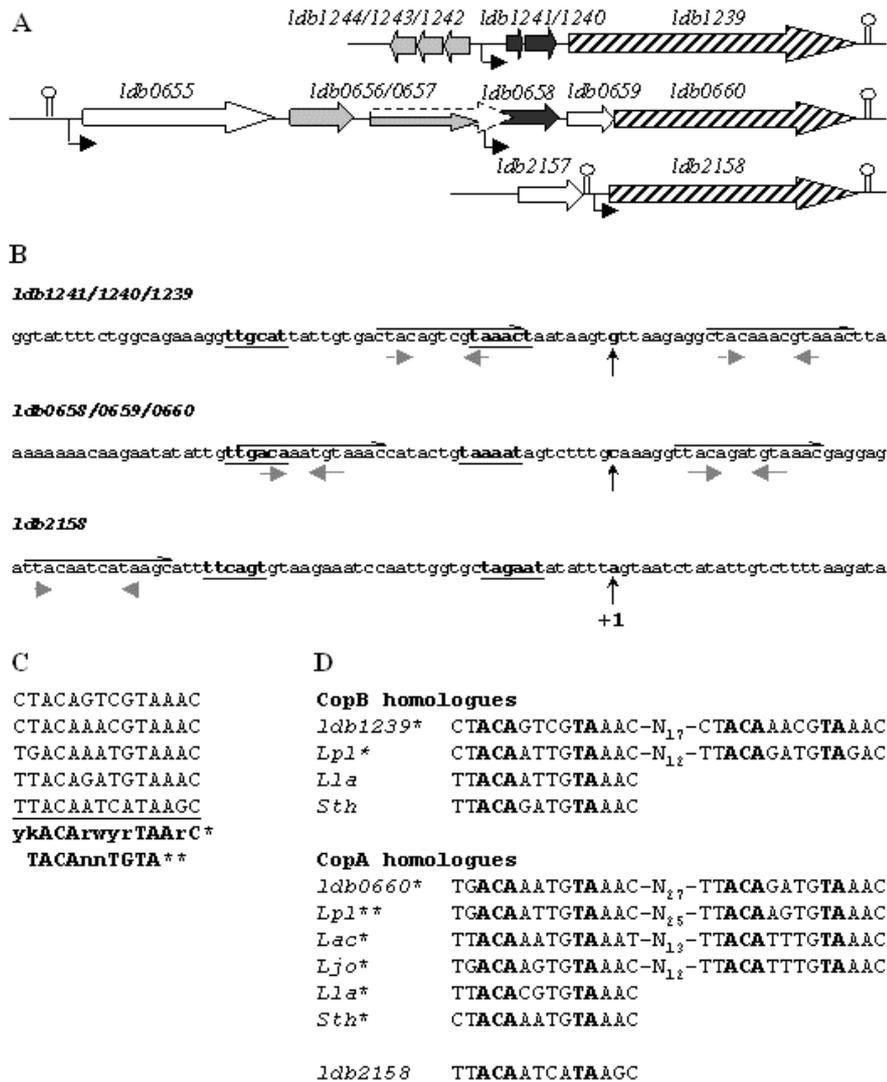


Fig. 4. Genetic organisation of CPx-ATPase loci.

A. CPx ATPase coding genes are indicated by hatched arrows, intact (*ldb0658*) or fragmented (*ldb1241/1240*) *copY* like regulator genes are indicated by black arrows. Transcription start sites are indicated by thin black arrows. Putative transcription terminators are indicated by stem and loop structures. *ldb0655* encodes a putative hydrolase, *ldb0659* encodes a conserved hypothetical protein, *ldb2157* encodes a conserved hypothetical protein. *ldb1242/1243/1244* constitute a 6-phospho-beta-glucosidase pseudogene, *ldb0656/657* constitute a penicillin binding protein (*pbp*) pseudogene. For *ldb0657*, the grey arrow indicates the part that is homologous to other *pbp* genes. At the site where these homologues contain a stop codon, the *ldb0657* open reading frame continues (dotted arrow), probably because gene degradation removed the stop codon. Likewise, gene degradation may have degraded a transcriptional terminator that may have been present downstream of the original *pbp* gene. **B.** Promoter regions upstream of the *ldb1241/1240-1239* and *ldb0658-0659-0660* operons and *ldb2158*. Transcription start sites are indicated as +1. Putative -35 and -10 promoter sequences are underlined. The common sequence motif detected by iMOMi is indicated by black arrows, and the inverted repeats within this motif by grey arrows. The motif is repeated twice in the promoter regions of the *ldb1241/1240-1239* and *ldb0658-0659-0660* operons. **C.** Alignment of the 5 motifs shown in part B. * consensus motif, **CopY binding site in Gram-positive bacteria (39). **D.** Alignment of the motifs shown in part B with motifs found

in promoter regions of homologous genes in a number of other LAB. *Lpl*, *Lactobacillus plantarum* WCFS1 (24); *Lac*, *Lactobacillus acidophilus* NCFM (3); *Ljo*, *Lactobacillus johnsonii* NCC 533 (40), *Lla*, *Lactococcus lactis* IL1403 (9); *Sth*, *Streptococcus thermophilus* LMG18311 (8). *L. acidophilus* and *L. johnsonii* do not encode CopB homologues. In the other bacteria mentioned, *copA* and *copB* make part of different transcription units, and are transcribed separately or cotranscribed with a *copY* homologue (*), or with a homologue of *ldb0659* (**). In the latter two cases the motifs shown are found upstream of the *copY* or *ldb0659* homologue, respectively. All bacteria mentioned except for *S. thermophilus* contain an *ldb2158* homologue, but the conserved motif is only found in the *L. bulgaricus* promoter region.

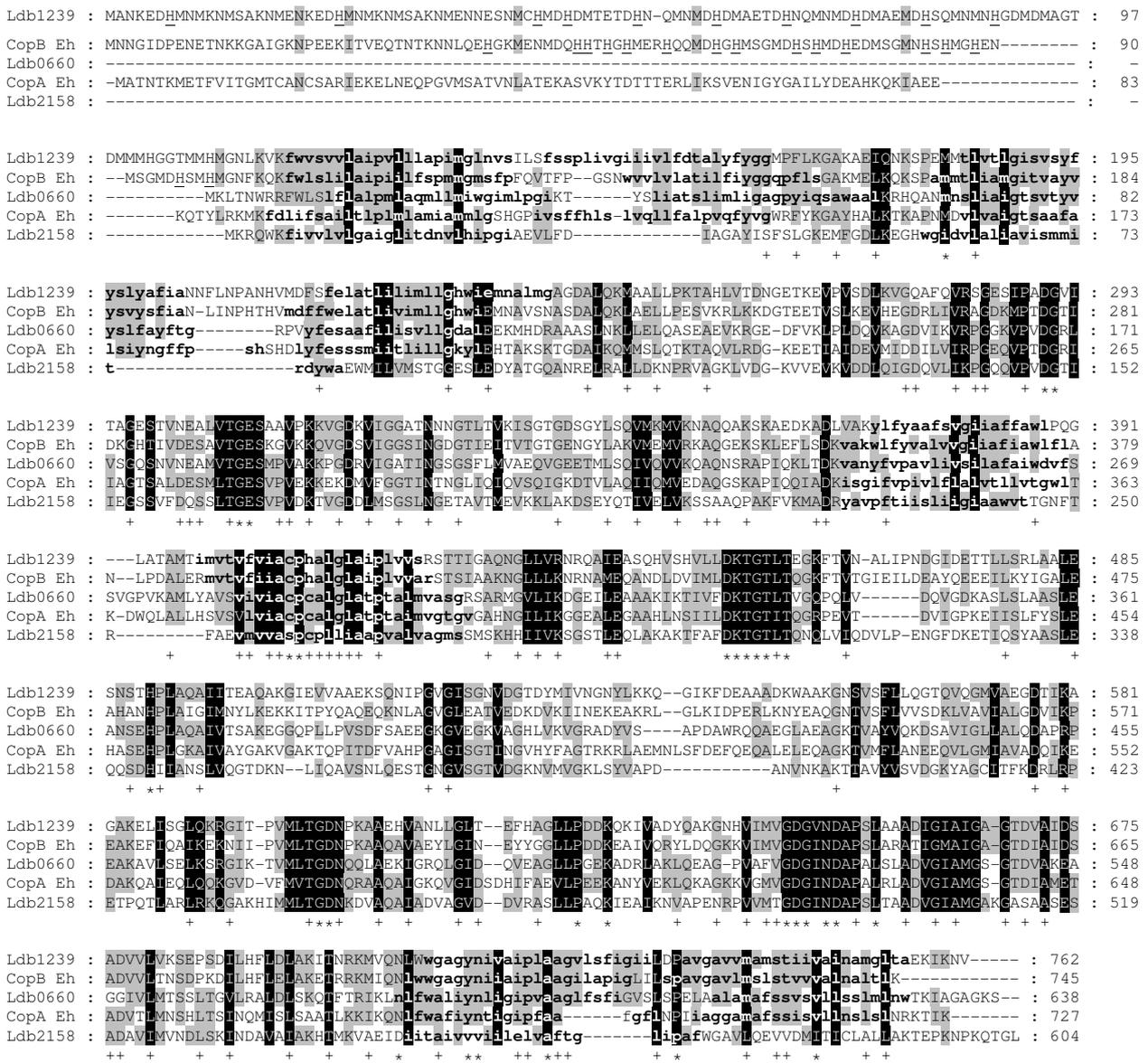


Fig. 5. Alignment of *L. bulgaricus* CPx-ATPases and Cu-ATPases from *E. hirae*.

Residues conserved or similar in all sequences are shaded in black, those conserved in at least three sequences are shaded in grey. Residues putatively involved in metal binding are underlined (CxxC motif or histidine rich region). Transmembrane domains detected by TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) are indicated in bold lower case. +, residue conserved in at least 70% of the probable copper transporting ATPases of the P-type ATPase Database (<http://www.patbase.kvl.dk/>). *, residues conserved in all of the probable copper transporting ATPases of the P-type ATPase Database. + and * refer to shaded residues.

2.2.3. Résultats complémentaires

2.2.3.1. Stress et adaptation croisés

Dans l'article précédant, des échantillons issus de cultures adaptées en présence d'acide lactique et d'HCl ont été utilisées (voir Table 4 p 179). Les stress physiologiques permettant de vérifier que les cellules étaient bien adaptées ont été réalisés en HCl. Nous avons préalablement vérifié que l'adaptation réalisée en présence d'acide lactique (acide faible) protège d'un stress réalisé en présence de HCl (acide fort) et inversement. Les résultats obtenus montrent qu'il y a bien une protection croisée entre les deux acides et que l'adaptation peut être visualisée par des stress effectués indifféremment en présence de l'un ou l'autre des acides (figure 19). Alors que le phénomène et le facteur d'adaptation sont reproductibles entre les expériences, les taux de survie obtenus avec un stress en présence d'acide lactique montrent une variation importante entre les expériences. Pour cette raison, nous avons décidé d'utiliser un stress en présence d'HCl pour les expériences complémentaires décrites dans l'article.

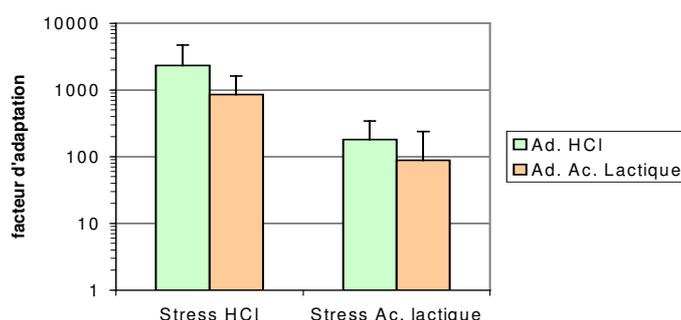


Figure 19 : Facteurs d'adaptation des stress croisés

Facteur d'adaptation = (taux de survie après adaptation)/(taux de survie avant adaptation).

Sont représentés les moyennes des facteurs d'adaptation calculés pour les cinq expériences ainsi que l'écart type.

Les adaptations sont réalisées pendant 90 min à pH 4,75 pour l'adaptation HCl et à pH 5,25 pour l'adaptation acide lactique. Les stress sont réalisés pendant 30 min à pH 3,6 pour le stress HCl et à pH 4,0 pour le stress acide lactique.

2.2.3.2. Mécanismes de régulation par atténuation

Au cours de l'étude, nous avons recherché en amont des gènes présentant une variation d'expression en condition d'adaptation à l'acide lactique, un motif de régulation conservé entre eux et entre les orthologues de chaque transporteur chez les lactobacilles dont nous avons la

séquence (*L. plantarum*, *L. johnsonii* et *L. acidophilus*). Cette recherche a été effectuée en utilisant le logiciel iMoMi (N. Pons *et al.*, en préparation). Outre le motif de CopY, nous avons pu mettre en évidence deux autres motifs. Ces deux motifs sont conservés chez les lactobacilles et se trouvent en amont de deux gènes dont les régions 5' non traduites sont particulièrement étendues (307 pb et 150 pb). Nous avons alors recherché si ces régions 5' non traduites présentaient un repliement potentiel particulier et s'il était conservé chez les autres lactobacilles.

Régulation de *ldb1810*

ldb1810 code un antiport Na^+/H^+ qui est réprimé d'un facteur 2 en condition d'adaptation à l'acide lactique. Ce gène présente une région 5' non traduite de 307 bases. Le motif conservé en amont de ce gène ressemble à la T-Box présente en amont de certains gènes de biosynthèse des acides aminés ou de tRNA synthétases chez les bactéries à Gram positif. Le système T-box est un mécanisme d'atténuation faisant intervenir un ARNt (Grundy et Henkin, 2002). La région 5' non traduite des gènes contrôlés par ce mécanisme présente deux structures mutuellement exclusives, une structure terminateur et une structure anti-terminateur. Lorsqu'il y a carence pour un acide aminé, les ARNt correspondant ne sont pas chargés. L'ARNt va alors se fixer sur la partie 5' non traduite de l'ARN messager en cours de synthèse et stabiliser une structure anti-terminateur par liaison d'une part de l'anticodon spécifique de l'acide aminé et d'autre part du bras accepteur de l'ARNt (séquence CCA). Cette stabilisation de la structure anti-terminateur va empêcher la structure terminateur de se former et ainsi permettre la transcription des gènes en aval. Lorsque l'acide aminé est présent en quantité suffisante, les ARNt spécifiques de cet acide aminé sont chargés et ils ne peuvent plus se fixer sur la région 5' non traduite. La structure terminateur se forme est la transcription des gènes est avortée.

Nous avons schématisé le repliement de la région 5' non traduite de *ldb1810* (figure 20). On retrouve bien une structure anti-terminateur (représentée) et une structure terminateur. Le codon correspondant à l'anticodon de l'ARNt est UUA, l'ARNt impliqué dans cette régulation serait donc l'ARNt^{Leu}. Cependant, on peut remarquer que dans la T-box, une partie seulement du bras accepteur de l'ARNt^{Leu} pourra se lier, car au lieu du motif UGG attendu, seuls sont conservés les bases UG. Chez les autres lactobacilles, on retrouve la séquence UGG. Il semble donc que chez les lactobacilles, l'expression de cet antiport Na^+/H^+ soit en relation avec la disponibilité en acide aminé et plus précisément avec la disponibilité en leucine pour *L. bulgaricus*.

C'est la première fois qu'un gène non relié au métabolisme des acides aminés est décrit comme ayant une régulation dépendante d'une T-Box. De plus, chez *L. plantarum*, l'homologue

de *ldb1810* est lui aussi réprimé en présence d'acide lactique ce qui renforce l'hypothèse d'une régulation de ce système par l'acidité (Pieterse *et al.*, 2005).

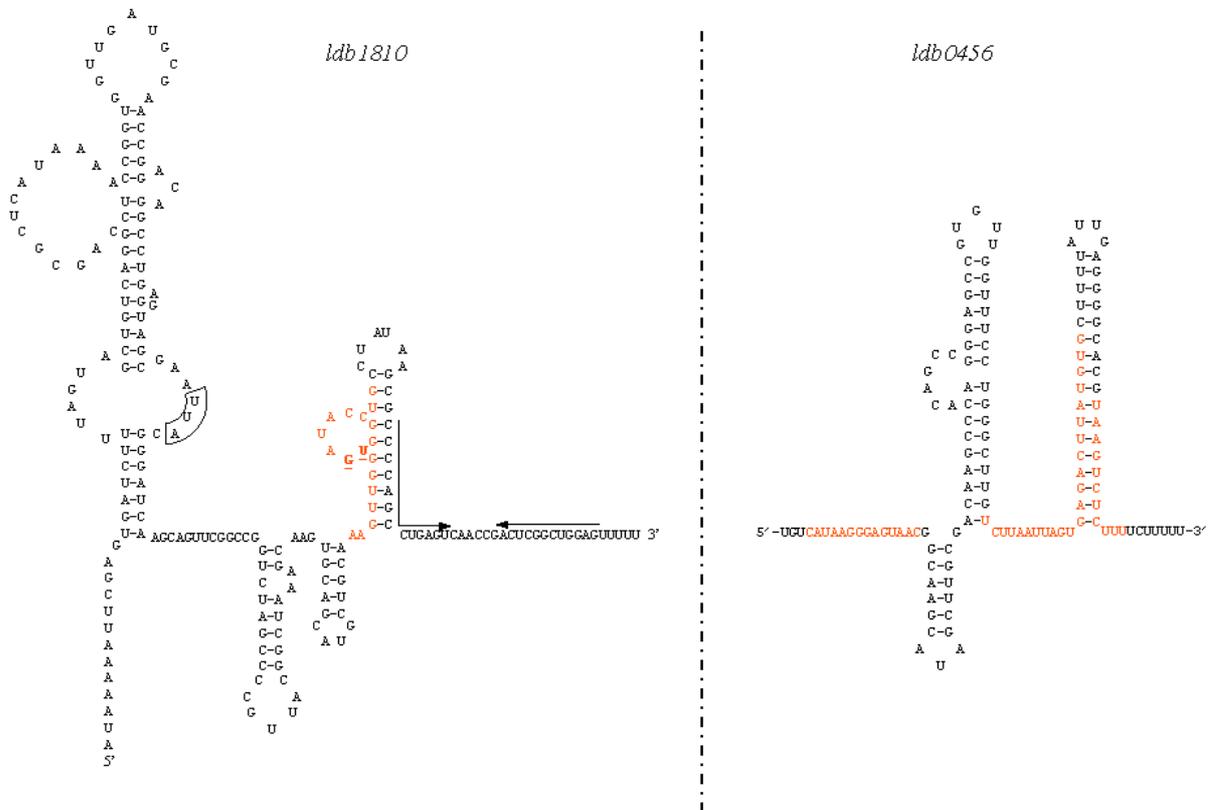


Figure 20 : Repléments des régions 5' non traduites en amont de *ldb1810* et *ldb0456*.

Pour *ldb1810*, le repliement est présenté avec la structure anti-terminateur. La région colorée en rouge correspond à la T-Box et les bases soulignées de la T-Box au site de liaison des bases CCA de l'ARNt. Les bases entourées correspondent au site de liaison de l'anticodon. Les flèches noires correspondent aux répétitions inversées formant le terminateur.

Pour *ldb0456*, le repliement est présenté avec la structure terminateur. Les motifs conservés avec *L. plantarum*, *L. johnsonii* et *L. acidophilus* sont colorés en rouge.

Régulation de *ldb0456*

Ldb0456 code pour une ATPase de type P dont la spécificité la plus probable est le transport de Ca^{2+} . Cette ATPase est réprimée d'un facteur proche de 2 en condition d'adaptation à l'acide lactique. Ce gène présente une région 5' non traduite de 150 bases. La recherche de motifs conservés entre les régions en amont du codon initiateur des orthologues de ce gène présents chez *L. plantarum*, *L. johnsonii* et *L. acidophilus* nous a permis d'identifier trois motifs conservés, la séquence de cette région, à l'exception des motifs, n'étant pas conservée. Nous avons pu proposer un repliement de la région 5' non traduite, conservée avec les trois autres lactobacilles et qui permet la formation d'une structure terminateur (figure 20). Ce repliement ne

fait pas intervenir de T-box. Il nous a été impossible de reconstituer une structure anti-terminateur à partir de ce repliement qui soit assez stable pour être significative. Les motifs conservés entre les lactobacilles ont certainement une importance dans la régulation du gène mais nous ne pouvons pas déterminer si ces motifs sont nécessaires à la fixation d'une protéine (qui lierait le calcium par exemple) ou d'une molécule chimique (qui pourrait être directement le calcium) qui permettraient la formation et/ou la stabilisation d'une structure anti-terminateur et ainsi la transcription de ce gène. Signalons que chez *L. plantarum*, l'homologue de ce gène a été décrit comme réprimé en réponse à l'acide lactique (Pieterse *et al.*, 2005).

2.3. Discussion

Chez *B. subtilis*, le développement de l'ATR est, au moins en partie, régulée par σ^B , qui induit aussi sa propre expression (Kovacs *et al.*, 1998, Volker *et al.*, 1999). *L. bulgaricus*, comme les autres BL, ne possède pas d'homologue de σ^B dans son génome et vu les résultats décrits dans cette partie, il est peu probable qu'un des facteurs σ de *L. bulgaricus* soit un analogue de σ^B . Le seul facteur σ dont l'expression varie durant l'ATR est réprimé d'un facteur 2,5, ce qui suggère qu'il ne serait pas impliqué dans la mise en place des mécanismes de l'ATR.

Les seules données sur la régulation de la tolérance au stress acide chez les BL disponibles dans la littérature portent sur un système à deux composants dont l'inactivation chez *L. acidophilus* (Azcarate-Peril *et al.*, 2005) et *L. lactis* (O'Connel-Motherway *et al.*, 2000) conduit à la diminution de la tolérance à l'acidité. Cependant chez *L. acidophilus*, le mutant est toujours capable d'induire une ATR. Alors que les résultats obtenus par Azcarate-Peril et coll. suggèrent que ce système à deux composants s'autorégule, aucune variation significative de l'expression de son homologue chez *L. bulgaricus* (*ldb1492-1493*) n'a été détectée.

Dans l'étude présentée, nous avons mis en évidence que 22% des régulateurs transcriptionnels d'ATCC 11842 (12/55) ont leur expression affectée par l'ATR. Vu les familles auxquels appartiennent ces régulateurs, notre hypothèse est que la majorité d'entre eux contrôle un régulon restreint, voire limité à l'opéron auquel ils appartiennent. Dans cette hypothèse, *L. bulgaricus* aurait alors un ensemble de régulons coordonnées par l'acidité du milieu plutôt que par un régulateur général. Cependant, l'analyse n'ayant été faite qu'au bout de 90 min d'adaptation, nous ne pouvons pas exclure qu'un régulateur général soit induit transitoirement au cours de l'adaptation et qu'à 90 min, son expression soit revenue à son niveau initial. Il serait intéressant de réaliser cette analyse à différents temps d'adaptation pour nous permettre de visualiser ce genre de profils en cinétique. De plus, l'activité d'un régulateur n'est pas toujours

corrélée avec la variation de son expression, car seuls les gènes soumis à auto-régulation ont une expression qui varie en fonction de leur activité. Il est donc possible qu'un certain nombre de régulateurs impliqués dans l'ATR n'ait pas été identifié par cette méthode.

Nous avons pu identifier une cible potentielle pour 10 régulateurs dont l'expression varie. Le régulateur codé par *ldb0347*, homologue à *purR*, régule probablement l'opéron de biosynthèse des purines. Chez *L. plantarum*, des études en transcriptomique ont montré que l'opéron *pur* est réprimé en présence d'acide lactique (Pieterse *et al.*, 2005), ce qui renforce les résultats obtenus pour ce régulateur. Le métabolisme des purines a déjà été étudié pour son rôle dans les réponses aux stress, notamment par l'obtention de mutants de cette voie plus résistants au stress acide (Rallu *et al.*, 2000) ou au stress oxydant (Fernandez *et al.*, 2004). La répression de cette voie serait donc cohérente avec ces observations.

Pour les 9 autres régulateurs, les cibles proposées codent pour des protéines impliquées dans des transports. Pour les deux régulateurs de la famille GntR, nous ne pouvons pas déterminer la spécificité des ABC transporteurs situés en aval de ces gènes. Pour *ldb2141*, *ldb2071* et *ldb1075* (et sa copie *ldb1132*), les transporteurs potentiellement régulés seraient impliqués dans l'efflux de composés toxiques pour la cellule. Nous avons confirmé que les cibles potentielles de Ldb2141 et Ldb2071, situées en opéron avec leur régulateur respectif, étaient bien induites en condition d'ATR. Il serait intéressant de déterminer la variation d'expression de la cible potentielle du régulateur dupliqué Ldb1075 (ou 1132), pour savoir si les résultats vont dans le même sens que pour les autres transporteurs de composés toxiques.

Enfin, pour quatre des régulateurs dont l'expression varie (*ldb0482*, *ldb0658*, *ldb0159* et *ldb2082*), nous avons identifié des cibles potentielles impliquées dans le transport d'ions et plus précisément dans le transport d'ions inorganiques pour trois d'entre elles, ce qui nous a conduit à nous intéresser aux transporteurs d'ions présents dans le génome de *L. bulgaricus*. L'analyse montre que pour 10 des 19 transporteurs d'ions identifiés, l'expression varie en condition d'adaptation à l'acidité. Un certain nombre de ces résultats est renforcé par les résultats de l'étude de Pieterse et coll. (2005) réalisée sur *L. plantarum*. En effet, chez cette bactérie, les homologues de *ldb0456* (ATPase de transport de cations), *ldb1810* (antiport Na^+/H^+) et l'opéron *pst* (transporteur de phosphate) sont eux aussi réprimés par l'acidité.

L'ATR modifie aussi l'expression de gènes potentiellement impliqués la régulation de réponses à d'autres stress comme le stress oxydant (*ldb0482*) ou les composés toxiques (*ldb1075*, *ldb1132*, *ldb2141* et *ldb2071*). Il serait intéressant de tester la résistance au stress oxydant des cellules adaptées à l'acidité pour déterminer si l'ATR protège d'un stress oxydant.

Plus globalement, les résultats obtenus dans cette étude nous montrent que nous n'avons probablement pas identifié l'ensemble des régulateurs de l'ATR puisque au moins quatre transporteurs d'ions sont régulés sans que nous puissions déterminer le régulateur incriminé. De plus, il est apparu que les mécanismes de régulation transcriptionnelle lors de l'ATR faisaient probablement intervenir des régulations par atténuation de la transcription en plus des régulateurs, ce qui, à notre connaissance n'avait pas encore été décrit.

3. Etude en milieu complexe

3.1. Introduction

Dans le chapitre précédant, l'étude transcriptomique a été limitée aux gènes codant des protéines impliquées dans deux types de fonction : les systèmes de régulation transcriptionnelle et les systèmes de transport d'ions. Une méthode largement utilisée chez les BL pour l'étude globale de l'ATR est l'étude protéomique par électrophorèse bidimensionnelle (2DE) (Budin-Verneuil *et al.*, 2005b ; Lim *et al.*, 2000 ; Len *et al.*, 2004 ; Wilkins *et al.*, 2002 et pour revue chez les lactobacilles : De Angelis et Gobbetti, 2004). La disponibilité de la séquence de *L. bulgaricus* ATCC 11842 permet de coupler la 2DE à l'identification des protéines par spectrométrie de masse. Nous avons donc opté pour cette méthode, disponible au laboratoire, pour identifier les protéines solubles de pI compris entre 4 et 7 pouvant avoir un rôle dans l'adaptation au stress acide d'ATCC 118412. L'étude de l'expression des gènes codant les protéines ainsi identifiées a été réalisée afin de déterminer si la variation d'abondance était liée à une régulation transcriptionnelle.

Cette étude a été réalisée en collaboration avec A. Fernandez et S. Boudebbouze. Ma contribution à ce travail a essentiellement porté sur la mise au point des conditions physiologiques d'adaptation et sur les expériences de PCR quantitatives.

3.2. Matériel et méthodes

Souche utilisée et conditions de culture. La souche utilisée dans cette étude est *L. delbrueckii* ssp. *bulgaricus* ATCC11842. *L. bulgaricus* est cultivé à 42°C en MRS (Difco). Les unités formant colonie (ufc) sont comptées après étalement sur milieu MRS gélosé et incubation à 42°C pendant 42 heures en condition anaérobie en jarre d'anaérobiose Anaeropack (Mitsubishi Gas Chemical Company) en présence de sachet AnaeroGen (Oxoid) contenant de l'acide ascorbique (absorbe l'oxygène et dégage du CO₂).

Détermination de la résistance à l'acidité et de l'ATR. Une colonie «fraîche» est utilisée pour ensemercer 1mL de MRS et est incubée 8 heures à 42°C sans agitation. La culture est alors diluée 10000 fois dans du MRS préchauffé à 42°C et la culture est incubée sur la nuit. La culture (DO₆₀₀ < 0.2) est ensuite diluée à DO₆₀₀ 0,05 dans du MRS à 42°C et incubée jusqu'à

DO₆₀₀ 0,2 (~2h). Pour mesurer la résistance à l'acidité, 1 mL de culture est centrifugé à 9000 × g (deux fois 2 min à température ambiante), et le culot est resuspendu dans 1mL de MRS ajusté à pH 3,8 (ou au pH de stress indiqué) avec de l'acide lactique (concentration finale ~0,17M). Immédiatement après resuspension et après 35 min à 42°C, 100µl de dilutions en séries en MRS sont étalés sur milieu MRS gélosé. Les ufc sont dénombrées après 36 h à 42°C. Le taux de survie est calculé en divisant le nombre d'ufc par mL après l'épreuve acide (35 min à pH 3,8), par le nombre d'ufc par mL immédiatement après resuspension à pH 3,8.

Pour étudier l'ATR, la culture à DO₆₀₀ 0,2 est divisée en deux échantillons. Pour l'échantillon non adapté, la résistance à l'acidité est évaluée par la survie à l'épreuve acide comme décrit ci-dessus. Pour l'échantillon adapté, 1 mL est centrifugé à 9000 × g (deux fois 2 min à température ambiante) et le culot est resuspendu dans 1 mL de MRS ajusté au pH d'adaptation (pH 4,9 ou au pH indiqué dans le texte) avec de l'acide lactique (48mM final) et incubée 40 min à 42°C. La résistance à l'acide lactique est ensuite mesurée comme décrit précédemment.

Extraction d'ARN et préparation des ADNc. Pour extraire les ARN, 10 mL de culture à DO₆₀₀ 0,2, avec ou sans traitement d'adaptation sont centrifugées à 11200 × g (6 min à température ambiante). Le culot est resuspendu dans 1 mL de tampon (10 mM Tris, 1 mM EDTA) préchauffé à 42°C et recentrifugé à 9000 × g (3 min). Le culot sec est ensuite congelé dans l'azote liquide et conservé à - 80°C. Les cellules sont décongelées à 4°C et resuspendues dans 400 µl de tampon (10 mM Tris, 1 mM EDTA) puis mélangées à 500 µL de phénol acide/chloroforme (5/1, v/v, Sigma). Les cellules sont lysées par agitation mécanique (3 x 45s, dans un appareil Bio 101 Fastprep) en présence de 0,6 g de billes de verre (diamètre <106 µm, Sigma). Les ARN sont extraits avec 1 mL de Trizol (Life Technologies) selon les recommandations du fournisseur. L'ADN est éliminé par traitement à la DNase (DNA-free system, Ambion) (0,04 U DNase 1 /µg ARN, 30 min à 37°C), et l'absence d'ADN est ensuite vérifiée par PCR. L'ARN est quantifié en mesurant l'absorbance à 260 nm. Les ADNc sont obtenus par transcription reverse de 1 µg d'ARN en utilisant la Powerscript Reverse Transcriptase (Clontech Laboratories) et 250 ng d'hexamères aléatoires (New England Biolabs) dans un volume de 20 µl, comme recommandé par le fournisseur.

PCR quantitative. Les oligonucléotides utilisés pour la PCR quantitative ont été dessinés avec le logiciel Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) et sont listés dans le tableau 13.

Tableau 13 : Oligonucléotides utilisés dans cette étude

Gène	Séquence des oligonucléotides (5'→3')	
	Direct	Inverse
<i>accC</i>	CTGGTTAAGGCCAGATCCT	GTTGATCCGGCATTTCGATAG
<i>clpL</i>	CACCACCCAGGATGAATACC	GACTCGGAGCGTTGACAGTT
<i>dnaK</i>	AAACGAAGTTGACGGCTTGAT	CTTCCAGGGCATCCTTAACCTG
<i>fabI</i>	ATTGCCGGAACAAGGATCT	CGCAGCAGTTACCCACTTCT
<i>fbA</i>	ACCCACGTAAGCTGCTCTTG	CAGCTTCCTTGTCGTCGATT
<i>folA</i>	CTGGCTGACCAGGTTGACA	CCAGTCAAGGTTTGGCATGT
<i>gmkI</i>	TAAGAGACCGGCTGGAACAC	TGTCGTTGACTACCGCGTAA
<i>groEL</i>	TCATCGCTGATGACGTGACT	CAAAGCCAGGAGCCTTAACA
<i>groES</i>	CTTGCTTCCAACGCTAAGGA	CTTGGACGGTCATAGGAAGC
<i>grpE</i>	GAGGCTTCCAGACAGCTGAA	CCGTCAGCCTTGATTTCTTC
<i>ptsH</i> (Hpr)	CGTCCAGCTACCTTGTTGGT	TGACATTACGCCCATGATTG
<i>ldb0095</i>	TGGTCCCTGCAGCATAACTT	TGCAGTAAGCGCAAAGTCAA
<i>ldb0262</i>	CTGAAGGCAACCCACAAGTT	TTTGAACCGCGCTTAATGTT
<i>ldh L</i>	CTGGTCATCTGCGACTTGAA	TGGTCTGGCCCGTAAAGTAG
<i>mvaC</i>	GCAGGTAGCCGATTACGACTT	TGCTTGTGACCTTGGCTAAA
<i>mvaS</i>	CACCTGCCTTACACCAAGATG	ACTTGGCTGAAAGCTGGAAGT
<i>oppF</i>	CAACAAGGACGTGCAGATGA	GGATCGGTTCTGCGATAATG
<i>pepD1</i>	CCTCCGTCACATCTTTGGTT	GTCGAAGTAGCGGTGGATGT
<i>pepP</i>	GCCTGGAAATCCATGAAATG	GTTCAACCGTGTGCACAATC
<i>pepV</i>	AGAGGAATTGATCGCCATTG	TTGGCAAAGCTGAGGAACCTT
<i>pgi</i>	CTGACAAGGAATTGCGTGAA	TTCTTGGCAGCCTTCTTGAT
<i>pox1</i>	GCAAGCAACGCTGGACTACT	GGCCAAGTGCCATACTTCAC
<i>ppaC</i>	CGGCGTTGACTACAAGGAAT	CGTCCAGGTCGATCAGTTCT
<i>pyk</i>	AGAAGGCGCAAGTTCACTA	CGTGGATCATGTCCTTGTG
<i>raiA</i>	AGTACAAGACCCGCGTCAAC	TCTTCAACCGGAGCATCTTC
<i>rodA</i>	AGACGATTCTGCGCCACTAC	GTAGGCGCTGTTGGAGTTGT
<i>sigA2</i>	GTCTGCTCTGGCCCTCAAGG	TCCGAACCGAGTCGTAAACCT

Les PCR sont réalisées dans un volume de 25 µl volume contenant 10 µl d'ADNc dilués 1000 fois, les oligonucléotides spécifiques (0,2 µM chacun) et 12,5 µL de ABsolute QPCR SYBR Green mix (Abgene). Le cycle de PCR est réalisé dans un appareil ABI 7700 (Applied Biosystems) avec les réglages suivants : 1 cycle à 95°C de 15 min, 40 cycles à 94°C pendant 15 s et à 60°C pendant 1 min. Les courbes de fusion servant à vérifier la spécificité de la réaction sont obtenues en augmentant la température de 1,5 °C par minute. L'efficacité de la PCR pour chaque couple d'oligonucléotides a été vérifiée en utilisant des dilutions en série d'ADNc ($1/180$, $1/540$, $1/1620$, $1/4860$). Le cycle auquel la fluorescence dépasse la valeur seuil détermine la valeur de cycle seuil (Ct). L'induction de l'expression des gènes lors de l'adaptation acide, est calculée avec la méthode comparative de cycles seuil (Livak et Schmittgen, 2001). Les équations suivantes sont utilisées :

- (i) $\Delta Ct = Ct$ du gène en condition adaptée - Ct du gène en condition non adaptée
- (ii) $\Delta\Delta Ct = \Delta Ct$ du gène - ΔCt du gène de référence (calibrateur)
- (iii) Niveau d'expression relatif = $2^{-\Delta\Delta Ct}$

Dans un premier temps, nous avons recherché un gène de référence. L'objectif était de trouver un gène dont l'expression ne varie pas entre les conditions utilisées. Différents gènes ont été testés : *pyk* (pyruvate kinase), *pgi* (glucose-6-P isomérase), *ldh* (L-lactate déhydrogénase), *pepV* (dipeptidase), *gmk1* (guanylate kinase), *folA* (dihydrofate déhydrogénase), *rodA* (rod-shaped determining protein) et *sigA2* (RNA polymérase sigma factor).

Pour vérifier que ces gènes étaient constitutivement exprimés, nous avons mesuré la quantité d'ADNc dans des échantillons issus d'au moins trois expériences indépendantes. Les échantillons d'ARN ont été prélevés à partir de cultures non adaptées et adaptées pendant 10, 20 et 40 min (figure 21).

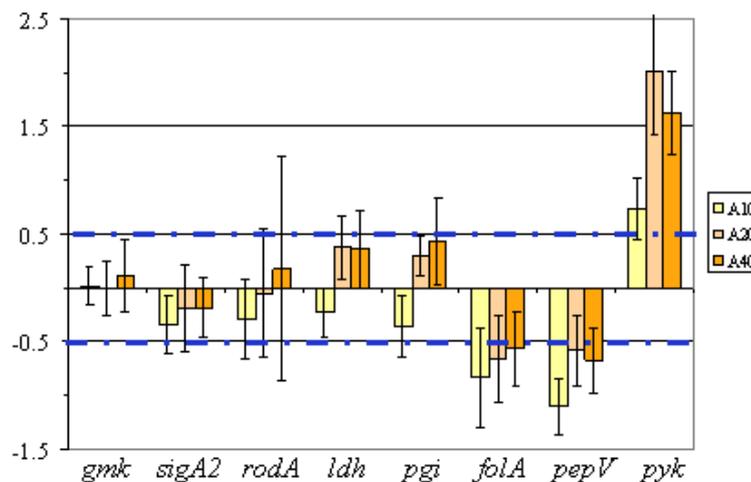


Figure 21 : Variation de l'expression des gènes testés comme gène de référence au cours de l'adaptation

Sur le graphique est représenté la moyenne et l'écart-type des ΔC_t mesurés sur au moins trois expériences indépendantes. A10, A20, A40 : mesures effectuées à 10, 20, 40 min d'adaptation. Les lignes bleues pointillées délimitent l'intervalle de 0,5 Ct.

Les gènes *pepV*, *folA* et *pyk* présentent des variations d'expression trop importantes pour être utilisés comme gène de référence. Parmi ceux ne variant pas de plus de 0,5 Ct entre la condition adaptée considérée et la condition non adaptée, une certaine variabilité a été observée pour les gènes *pgi*, *ldh*, *rodA* et *sigA2*. Le gène *gmk1* présentait les plus faibles variations d'expression, comprises entre 1,4 fois et 0,7 fois (correspondant à 0,5 Ct) dans toutes les conditions testées et a été retenu comme gène de référence.

Extraction des protéines. Les protéines sont extraites d'après Lim *et al* (2000). Les culots bactériens, prélevés à DO_{600} 0,2, avec ou sans adaptation, sont lavés dans un tampon (Tris 50 mM, pH 8 à 4°C) et resuspendus dans un tampon de lyse (Urée 7 M, Thiourée 2 M, Chaps

4%, Triton X100 0,25%, IPG buffer 0,5%, Bleu de bromophénol, $MgCl_2$ 1 mM, PMSF 1 mM, DTT 1 mM, Rnase 4 $\mu g/ml$, Snase 0,2 U/ml). Les cellules sont lysées mécaniquement au Cell-Disrupter (Constant) (2,7 Kbar) et le surnageant, récupéré après centrifugation (20 min à 13000 rpm et à 4°C), est conservé à - 80°C. La quantité de protéines est mesurée par un test de Bradford.

Electrophorèse bidimensionnelle. La migration en première dimension est effectuée en chargeant 300 μg de protéines par strips (Immobilin drystrips pH 4-7, Amersham) et la migration est effectuée pendant 18h30. L'équilibration des strips est ensuite effectuée en deux étapes : une incubation de 12 min dans 10 mL de solution d'équilibration-DTT (Urée 6 M, bleu de bromophénol, SDS 2%, Tris 50 mM pH 8,8, glycérol 30%, DTT 1%) suivie d'une incubation de 12 minutes dans 10 ml de solution d'équilibration-Iodoacétamide (Urée 6 M, bleu de bromophénol, SDS 2%, Tris 50 mM pH 8,8, glycérol 30%, Iodoacétamide 5%). La migration en deuxième dimension est effectuée en SDS-PAGE contenant 12% d'acrylamide pendant 16 h 30 à 10 mA par gel. Les gels sont colorés dans le bleu de Coomassie Bio-Safe (Biorad) pendant 1 h puis décolorés dans l'eau milliQ. Les gels sont enfin numérisés avec un scanner EPSON Expression 1640XL calibré pour que l'intensité soit intégrée de façon linéaire. Les analyses quantitatives et comparatives sont effectuées avec le logiciel Phoretix 2D Expression (Non Linear Dynamics). Trois électrophorèses 2D effectuées avec des extraits protéiques indépendants obtenus à partir de cultures adaptées sont analysées et comparées à celles obtenues pour la condition non adaptée (condition de référence). Un facteur de variation est calculé pour chaque spot : il est égal au rapport entre l'intensité d'un spot dans la condition adaptée et l'intensité du spot équivalent dans la condition non adaptée. Les valeurs positives correspondent à des spots plus abondants après adaptation qu'avant ; les valeurs négatives correspondent aux spots sous-représentés dans la condition adaptée par rapport à la condition non adaptée. Un t-test (loi de Student) est réalisé pour chaque spot dont l'abondance varie entre les deux conditions testées pour s'assurer de la validité statistique de la mesure.

Identification des spots. Après analyse comparative des profils 2D, les spots variants sont découpés, lavés en acétonitrile-hydrogencarbonate d'ammonium 50 mM (1/1, v :v) pendant 15 min en agitant, puis les extraits sont séchés 30 min au Speed Vac. La trypsinolyse est effectuée après réhydratation dans 20 μL d'une solution contenant 0,25 μg de trypsine (Promega) diluée dans de l'hydrogencarbonate d'ammonium (50 mM). Le mélange de digestion est incubé 16 à 18 h à 37°C. La réaction est arrêtée par ajout de 2 μL d'acide trifluoroacétique 5 %. L'analyse peptidique est effectuée par MALDI-TOF (Matrix Assisted laser Desorption Ionisation – Time Of Flight) avec une matrice d'alpha cyano-4-hydroxy-cinnamic acid (5 mg/mL). Les résultats

sont analysés à l'aide du logiciel Voyager DE super STR MALDI-TOF mass Spectrometer (Applied Biosystems) calibré sur les pics des spectres peptidiques d'autolyse de la trypsine. L'identification des protéines est effectuée grâce au logiciel Protein Prospector en utilisant la base de données de séquences de *L. bulgaricus* ATCC 11842.

3.3. Résultats

3.3.1. Résistance à l'acidité en MRS

Pour déterminer les conditions de stress pour l'étude de l'adaptation en MRS, nous avons soumis des cultures de *L. bulgaricus* en phase exponentielle de croissance à des épreuves acides à différents pH et pendant des temps différents (figure 22). Comme attendu, la survie de la

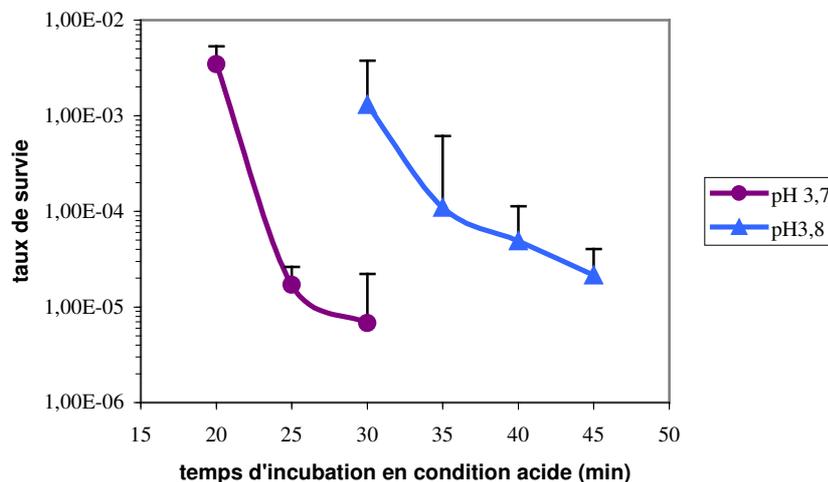


Figure 22 : Taux de survie de *L. bulgaricus* en fonction du temps et du pH d'incubation en présence d'acide lactique.

Taux de survie = (nombre d'ufc après épreuve acide)/(nb d'ufc avant l'épreuve).

Sont présentés la moyenne et l'écart-type pour 3 expériences indépendantes

culture diminue fortement en fonction du temps et du pH. Comme observé précédemment pour d'autres BL, une assez faible différence de durée ou de pH d'incubation en condition acide peut conduire à une modification très sensible de la survie. A titre d'exemple une incubation à pH 3,7 de 20 min ou 25 min se traduit par une différence de survie d'un facteur 100. De façon similaire, 30 min d'incubation à pH 3,8 ou 3,7 aboutit à une différence de survie d'un facteur 100. Pour l'étude de l'adaptation, nous souhaitons nous placer à des taux de survie pour des cultures non

adaptées proche de 10^{-4} . Pour l'ensemble de l'étude en MRS, nous avons utilisé une épreuve acide de 35 min à pH 3,8, ce pH étant obtenu par addition d'acide lactique.

3.3.2. Adaptation

Pour identifier les conditions d'adaptation optimales de la souche en MRS, nous avons soumis la culture à une incubation de 40 min (soit environ un temps de génération en MRS) à des pH intermédiaires pendant différents temps (figure 23). Une incubation de 40 min à des pH compris entre 4,3 et 5,5 augmente la survie de la culture à l'épreuve acide d'un facteur supérieur à 100. Comme attendu, *L. bulgaricus* développe une ATR en MRS. Le meilleur taux de survie est obtenu avec une pré-incubation de 40 min à pH 4,9, la survie est alors améliorée d'un facteur supérieur à 750. C'est cette condition que nous utiliserons dans la suite de l'étude.

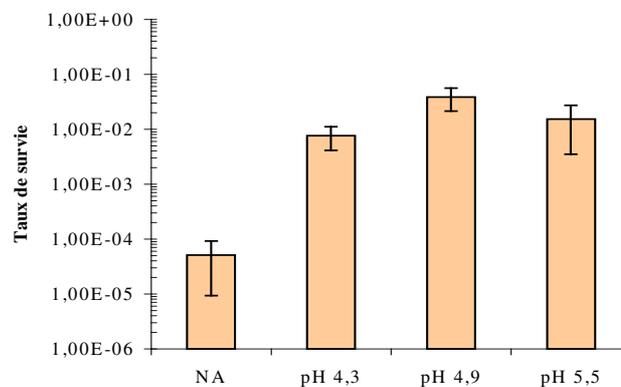


Figure 23 : Taux de survie en fonction du pH d'adaptation.

Pour chaque condition, les résultats représentent la moyenne et l'écart-type de trois expériences indépendantes.

NA : non adapté ; pH x : après adaptation à pH x pendant 40 min

3.3.3. Protéomique

Pour identifier les protéines impliquées dans le mécanisme d'adaptation, une analyse en électrophorèse 2D a été effectuée. Les protéines ont été extraites à partir de trois cultures indépendantes non adaptées et adaptées (40 min à pH 4,9). L'adaptation a été vérifiée physiologiquement par la détermination des taux de survie des cultures adaptées et non adaptées. Dans les trois cas, le facteur d'adaptation est supérieur à 100 (résultats non montrés).

En moyenne, 280 spots sont détectés sur les gels 2D colorés au bleu de Coomassie (figure 24). La comparaison des spots détectés sur les gels a permis d'identifier 21 spots dont l'intensité

varie d'un facteur supérieur à 1,75 (T-test inférieur à 0,03) entre les échantillons non adaptés et adaptés.

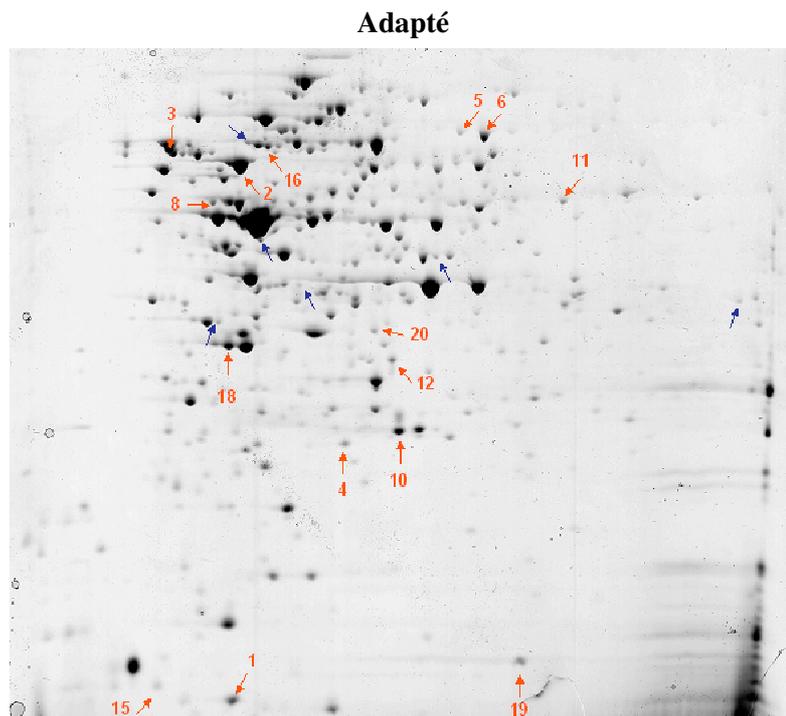
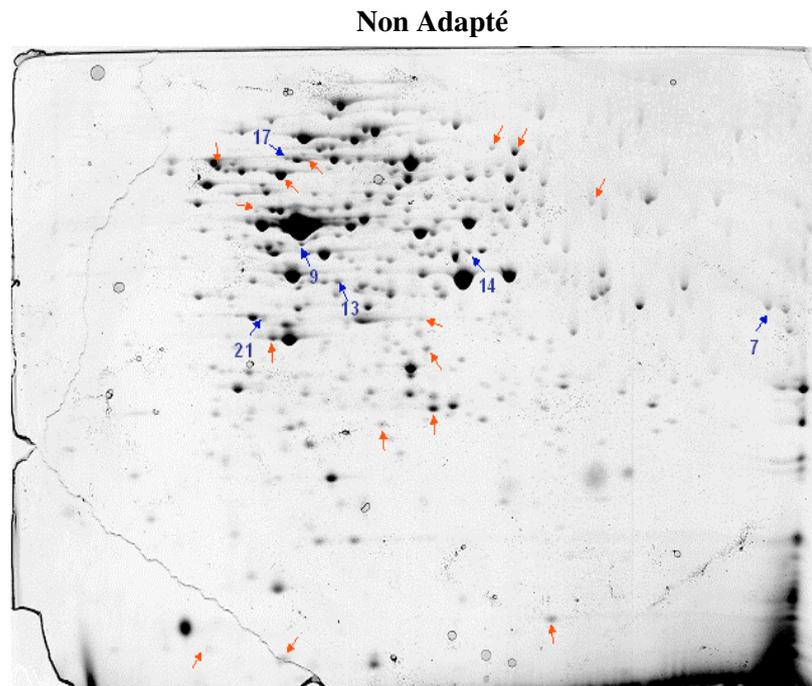


Figure 24 : Exemples de protéomes de *L. bulgaricus* ATCC 11842 avant (non adapté) et après adaptation 40 min à pH 4,9.

Les spots indiqués en rouge sont plus abondants après adaptation, alors que ceux indiqués en bleu sont réprimés après adaptation. Les numéros de spots indiqués sont ceux utilisés dans le tableau 14.

Ces 21 spots ont été identifiés par une analyse en spectrométrie de masse MALDI-TOF. Les protéines identifiées ont été classées en cinq groupes fonctionnels (tableau 14).

Le premier groupe comprend les protéines chaperons GroES, GroEL, DnaK et GrpE, ainsi que deux isoformes de ClpL, toutes induites après adaptation. Le deuxième groupe comprend des protéines impliquées dans le métabolisme carboné. HPr est un composant du système de

transport de sucres PTS («PhosphoTransferase System») et Fba est une enzyme de la glycolyse, la fructose bisphosphate aldolase. Ces deux protéines sont accumulées après adaptation. Deux isoformes de la pyruvate oxydase sont détectées, la forme de pI le plus élevé (~ 5) augmente en condition d'adaptation alors que l'autre forme (pI ~ 4,96) diminue. Le troisième groupe comprend des protéines impliquées dans le métabolisme des lipides. Le niveau d'expression de AccC et de FabI augmente d'un facteur similaire (2,6). Ces protéines sont toutes deux impliquées dans la synthèse de type II des acides gras (Marrakchi *et al.*, 2002) et sont codées par des gènes qui semblent faire partie d'un même opéron. D'autre part, MvaC et MvaS sont moins abondantes en condition d'adaptation à l'acide lactique. Ces deux enzymes sont impliquées dans les premières étapes de la voie du mevalonate de la synthèse d'isoprenoides. Le quatrième groupe comprend les protéines impliquées dans le transport et la dégradation de peptides. Une des protéines, PepD1 est induite alors que les deux autres, PepP et OppF sont en quantité moins abondante après adaptation.

Tableau 14 : Protéines induites ou réprimées lors de l'adaptation à l'acide lactique

N° de spot	Facteur d'expression A40/NA	T-test	pI th	pM th, (kDa)	pI obs sur gel	pM obs sur gel	n° ldb	Protéine	Fonction	CAI
Protéines chaperons										
1	4,3	0,002	4,83	10,3	4,9	9	Ldb1618	GroES	Chaperone protein 10kDa	0,529
2	2,2	0,000	4,86	57,3	4,9	60	Ldb1617	GroEL	Chaperone protein 60kDa	0,590
3	2,2	0,001	4,68	66,1	4,7	65	Ldb1313	DnaK	Chaperone protein	0,657
4	1,8	0,007	5,42	20,6	5,3	20	Ldb1314	GrpE	Chaperone protein	0,446
5	3,6	0,028	5,59	76,9	5,5	70	Ldb2102	ClpL	Clp protease, ATP-binding subunit	0,455
6	2,1	0,001	5,59	76,9	5,6	70	Ldb2102	ClpL	Clp protease, ATP-binding subunit	0,455
Métabolisme carboné										
Glycolyse										
18	1,9	0,000	4,88	33,1	4,9	28	Ldb1544	Fba	Fructose bisphosphate aldolase	0,813
Transport										
15	2,4		4,8	9,2	4,7	10	Ldb0576	HPr	Phosphocarrier protein HPr	0,804
Métabolisme du pyruvate										
16	2,5	0,026	4,96	67,7	5,05	65	Ldb2213	Pox1	Pyruvate oxidase	0,763
17	-1,9	0,021	4,96	67,7	4,9	70	Ldb2213	Pox1	Pyruvate oxidase	0,763
Métabolisme des lipides										
14	-1,9	0,001	5,43	43,0	5,4	41	Ldb0881	MvaS	Hydroxymethylglutaryl-coenzyme A synthase	0,463
13	-6,9	0,000	4,98	38,5	5,1	35	Ldb0879	MvaC	Acetyl-CoA acetyltransferase	0,445
12	2,6	0,003	5,23	26,8	5,25	26	Ldb0910	FabI	Enoyl reductase	0,385
11	2,6	0,024	5,74	50,8	5,8	55	Ldb0907	AccC	Acetyl-CoA carboxylase subunit	0,363
Dégradation et transport des peptides										
8	4,5	0,016	4,81	53,5	4,8	55	Ldb0511	PepD1	Dipeptidase	0,487
9	-1,9	0,026	4,92	40,3	4,95	44	Ldb1429	PepP	X-Pro aminopeptidase	0,439
7	-3,4	0,001	6,39	36,1	6,9	32	Ldb1385	OppF	Oligopeptide ABC transporter, ATP-binding protein	0,359
Divers										
19	1,9	0,019	5,59	13,3	5,6	12	Ldb0262		Conserved hypothetical protein	0,623
21	-2,1	0,017	4,75	34,3	4,85	31	Ldb0995	PpaC	Manganese-dependent pyrophosphatase	0,611
10	1,8	0,004	5,32	21,9	5,28	22	Ldb0606	RaiA	Ribosome-associated inhibitor	0,610
20	8,0	0,024	5,27	31,0	5,23	29	Ldb0095		Oxidoreductase	0,414

Les protéines dont la quantité sur gel diminue après adaptation sont grisées.

pI th : point isoélectrique théorique. PM th : poids moléculaire théorique. Obs : observé. CAI : Codon Adaptation Index.

Le dernier groupe comprend des protéines de fonctions diverses dont une oxydoréductase de la famille aldo/keto réductase qui présente le taux d'induction le plus élevé (facteur 8). Ce groupe contient aussi une protéine associée aux ribosomes, RaiA, homologue à RaiA de *E. coli* qui inhibe la traduction en se fixant aux ribosomes en condition de stress froid dont l'abondance augmente lors de l'adaptation (Wilson et Nierhaus, 2004).

Les CAI des protéines identifiées dans cette étude sont compris entre 0,359 et 0,813. La présence de protéines dont le CAI est inférieur à 0,450 (seuil pour la prédiction des protéines fortement exprimées) indique qu'un maximum de protéines présentes sur le gel ont été analysées, même celles présentant une abondance moyenne voire faible (figure 24).

3.3.4. Etude transcriptionnelle

Pour déterminer si les variations protéiques observées entre les conditions adaptées et non adaptées sont liées à une régulation transcriptionnelle ou post-transcriptionnelle, nous nous sommes intéressés à l'expression des gènes codant les protéines d'intérêt identifiées en 2DE. Des ARN totaux ont été préparés en condition non adaptée et après 10, 20 et 40 min d'incubation à pH 4,9 (adaptation). Nous avons choisi un suivi cinétique de l'expression des gènes car la plus forte abondance d'une protéine à 40 min d'adaptation pourrait résulter d'une induction transitoire, plus ou moins précoce du gène correspondant.

Sur les 19 gènes dont l'expression est suivie pendant l'ATR, 11 gènes présentent une expression modifiée de plus d'un facteur 2 pendant l'ATR, 6 gènes s'expriment de façon relativement constante (variation inférieure à un facteur 2) et 2 gènes présentent des fluctuations de l'expression variables suivant les expériences (figure 25).

Les variations d'expression *groES* et *groEL* sont très proches, ce qui est cohérent avec la structure opéronique des ces deux gènes. La même observation est valable pour les gènes *dnaK* et *grpE* ainsi que pour *fabI* et *accC*. Les profils d'induction des gènes, *groES-groEL* et *dnaK-grpE* sont relativement comparables : l'induction est importante dès 10 min d'adaptation, atteint son maximum à 20 min d'adaptation (facteur d'induction de 40 pour *groES* et *groEL* et de 17 pour *dnaK* et *grpE*) alors qu'après 40 min d'adaptation, les gènes sont plus exprimés que dans l'échantillon non adapté mais les facteurs d'induction ont fortement diminué (~16 fois pour *groES* et 4 fois pour *dnaK-grpE*). Ces résultats sont cohérents avec une régulation coordonnée de ces deux opérons. *A contrario*, l'induction du gène *clpL* au niveau protéique, n'est pas confirmé au niveau transcriptionnel, les résultats étant non reproductibles entre les différentes expériences.

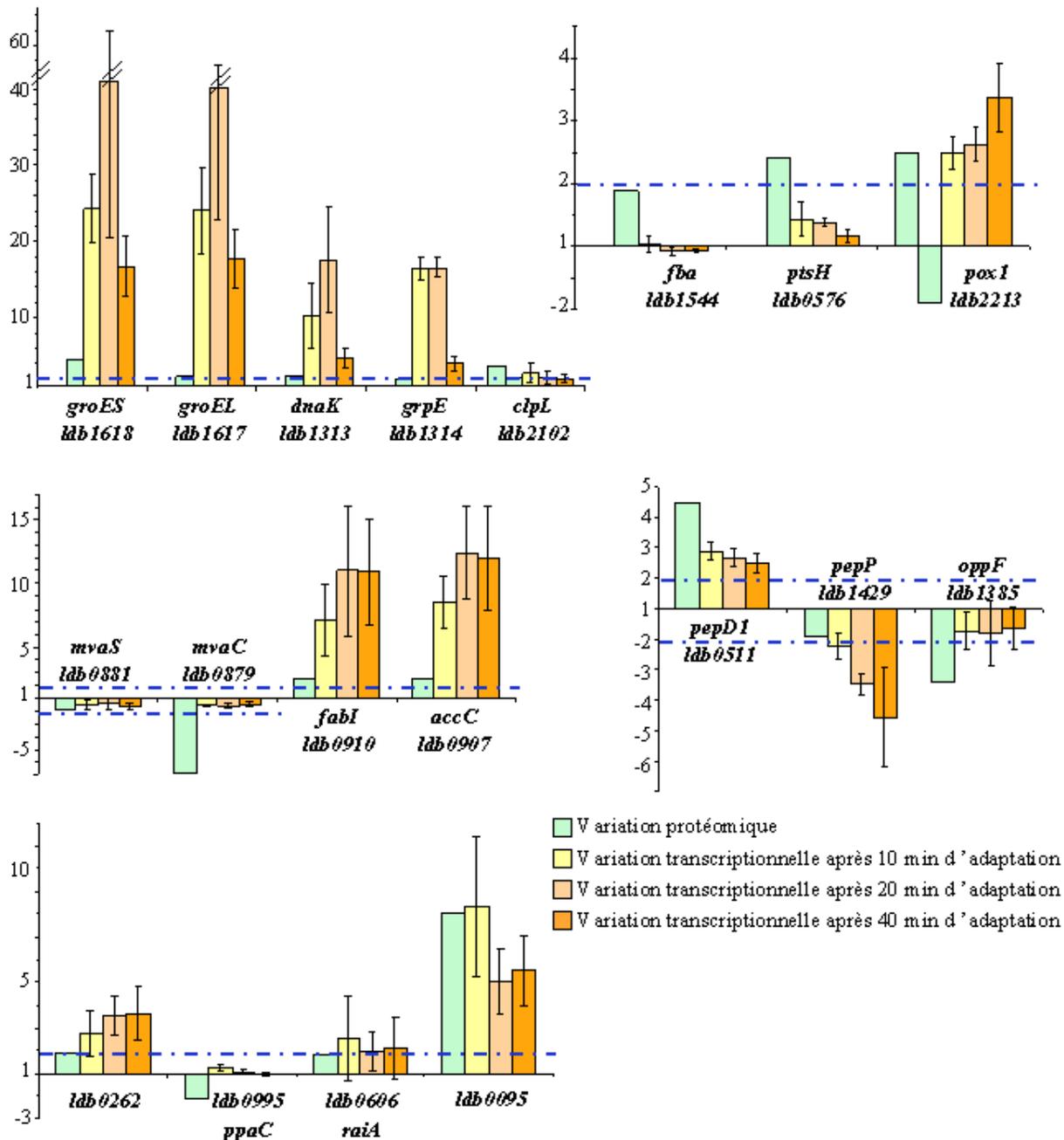


Figure 25 : Variations transcriptionnelles des gènes codant les protéines affectées par l'adaptation à l'acide lactique.

Les gènes sont groupés par classe fonctionnelle, comme défini au tableau 14. Sont représentés les moyennes et écarts-type de trois expériences indépendantes. Les valeurs positives signifient une induction du gène, les valeurs négatives, une répression. Les lignes bleues pointillées représentent un facteur 2 en induction ou en répression. Dans le cas de deux isoformes détectées lors de l'étude en protéomique, deux histogrammes verts sont représentés pour le même gène.

Les résultats obtenus pour les gènes codant les protéines impliquées dans la dégradation et le transport des peptides semblent indiquer que *pepD1* et *pepP* sont régulés au niveau

transcriptionnel. Les résultats obtenus pour *oppF* ne permettent pas de mettre en évidence une variation de la quantité des transcrits significative entre les échantillons adaptés et non adaptés.

L'induction transcriptionnelle des gènes impliqués dans la synthèse des acides gras, *accC* et *fabI*, est supérieure à un facteur 10 et elle est relativement constante au cours de l'adaptation. L'expression des gènes *mvaC* et *mvaS* n'est pas significativement réprimée en condition d'ATR, la diminution de l'abondance des protéines codées par ces gènes pourrait résulter d'une régulation post-transcriptionnelle. La protéine Hpr pourrait elle aussi être la cible d'une régulation post-transcriptionnelle puisque l'expression du gène correspondant, *ldb0576* varie de façon non significative au cours de l'ATR. L'expression des gènes *fba* et *ppaC* n'est pas affectée en condition d'adaptation.

Les autres gènes induits en condition d'adaptation sont : le gène *ldb2213* codant la protéine Pox1, *ldb0262* et *ldb0095*. Le produit de traduction du gène *ldb0095* est la protéine la plus affectée par l'adaptation acide suivant l'analyse en 2DE. Malheureusement nous ne pouvons prédire qu'une fonction générale, elle appartient à la famille des aldo/keto réductases. Le cas du gène *ldb2213* est intéressant puisqu'une induction transcriptionnelle est confirmée alors que l'analyse 2D révélait deux isoformes, une qui s'accumulait alors que l'autre diminuait. L'analyse des transcrits révèle donc qu'il ne s'agit pas uniquement d'une modification post-traductionnelle de la protéine Pox mais qu'il y a bien induction du gène.

3.4. Discussion

3.4.1. ATR et stimulon de réponse au choc thermique

L'induction des protéines chaperons lors de l'adaptation au stress acide avaient déjà été mise en évidence par protéomique 2D chez plusieurs bactéries lactiques, et notamment dans une autre souche de *L. bulgaricus*, Vi1010, où les protéines GroES, GroEL et DnaK avaient été identifiées par séquençage N-terminal (Lim *et al.*, 2000 ; Guillouard *et al.*, 2004). Dans cette étude, menée sur ATCC 11842, nous confirmons l'induction de ces trois protéines et nous démontrons qu'il s'agit d'une régulation transcriptionnelle de l'opéron *groESL* et des gènes *dnaK* et *grpE*. Les gènes *dnaK* et *grpE* appartenant à l'opéron *hrcA/grpE/dnaK/dnaJ*, il est probable que les gènes *hrcA* et *dnaJ* sont eux aussi induits transcriptionnellement dans ces conditions. La régulation de ces opérons est classiquement effectuée par le répresseur HrcA, mais nous ne pouvons pas affirmer que l'induction de ces gènes est bien dépendante de l'inactivation de ce répresseur dans les conditions testées.

Un élément de réponse pourrait être apporté par l'étude transcriptionnelle des autres membres du régulon HrcA défini par analyse *in silico* (cf Résultats §3.5.1.1). L'induction de l'ensemble du régulon HrcA lors de l'ATR conforterait l'hypothèse d'une régulation médiée par HrcA en réponse à l'acidité du milieu. Cependant, parmi les produits de ces gènes, ClpP devrait être détecté par les analyses protéomiques, elle fait partie des protéines de CAI élevé (0,555) et de pI et PM compatibles avec les conditions 2DE utilisées mais il n'a pas été identifié sur le gel de référence d'ATCC 11842 (S. Boudebouze, com. pers.). L'induction des protéines de choc thermique lors de l'adaptation acide expliquerait l'observation que l'ATR induit une résistance accrue au choc thermique (EM Lim, com. pers.).

L'augmentation de deux formes de la protéine ClpL est aussi observée à 40 min d'adaptation. ClpL est impliquée dans la résistance au stress thermique chez *S. thermophilus* (Varcamonti *et al.*, 2006). Par ailleurs, l'induction de ClpL lors de l'ATR a déjà été décrit chez *S. mutans* (Wilkins *et al.*, 2002 ; Len *et al.*, 2004) et chez *S. pneumoniae* (Martin-Galiano *et al.*, 2005). Cependant, chez *L. bulgaricus*, *clpL* ne ferait pas partie du régulon HrcA, contrairement à ce qui est décrit chez *L. rhamnosus* (Suokko *et al.*, 2005). Le résultat de l'étude transcriptionnelle de *clpL* est ambigu : le facteur d'induction est faible (entre 2,6 et 2) et les écarts-type obtenus sont élevés (de 1,26 à 0,53), ce qui nous ne permet pas de conclure si *clpL* est induit. Cette situation est très différente de celle rencontrée pour les gènes *groESL*, *dnaK* et *grpE*, pour lesquels l'induction transcriptionnelle est importante et indéniable. Cette différence est en accord avec deux modes de régulation différents pour *clpL* et le régulon HrcA. Il serait intéressant de déterminer la variation d'expression lors de l'ATR des gènes appartenant au stimulon choc thermique mais qui ne font pas partie du régulon HrcA. Dans le cas d'une induction transcriptionnelle, se poserait la question d'une régulation coordonnée et du régulateur impliqué.

3.4.2. ATR et métabolisme carboné

Bien que les gènes codant Hpr et CcpA soient détectés dans le génome d'ATCC 11842, il existe peu d'évidences de l'existence d'une répression catabolique. En effet, celle-ci a pour but de favoriser les sucres transportés par les systèmes PTS et notamment le glucose, considérée comme la source de carbone la plus énergétique. Le lactose est transporté par une perméase et ATCC 11842 possède seulement deux systèmes PTS complets. Or, en milieu chimiquement défini, *L. bulgaricus*, se développe mieux en lactose qu'en glucose (Chervaux *et al.*, 2000). Lorsque les deux sources de carbone sont présentes dans le milieu, *L. bulgaricus* ne présente pas de phase diauxique, ce qui suggérerait que les deux sucres sont consommés en même temps.

Le gène codant la protéine responsable de la phosphorylation de Hpr, l'Hpr kinase, est détecté dans le génome de *L. bulgaricus*. On pourrait donc s'attendre à détecter sur gel deux isoformes de Hpr selon le degré de phosphorylation de la protéine. Cette étude ne nous a permis de détecter qu'une seule isoforme et ne nous ne pouvons pas prédire de quelle forme il s'agit. L'augmentation de l'abondance de la protéine Hpr semble être post-traductionnelle car aucune induction n'a pu être mise en évidence au niveau transcriptionnel. La protéine Hpr a déjà été associée à l'ATR. En effet, une étude en 2DE menée chez *L. lactis* montre que la protéine Hpr est réprimée pendant l'ATR (Frees *et al.*, 2003). Cependant, dans cette étude, l'isoforme détectée n'a pas été caractérisée. Chez *L. casei*, un mutant du site de phosphorylation de Hpr par la Hpr kinase/phosphorylase (dans ce cas, Hpr ne peut plus être phosphorylé par cette enzyme) présente une sensibilité accrue au stress acide, alors que le mutant de l'Hpr kinase ne présente pas de phénotype (thèse de S. Beaufils).

Des sites Cre (pour Catabolite responsive element) putatifs, site de liaison du régulateur transcriptionnel CcpA responsable de la répression catabolique, sont détectés en amont de différents gènes associés au métabolisme carboné et notamment dans les promoteurs de *pfk*, *pyk*, *eno*, *gpmA*, codant des enzymes de la glycolyse, *ptsH* et *ptsI* codant Hpr et HprK, *ccpA* et en amont d'un gène codant une sous-unité IIA du système PTS β - glucoside. L'absence de régulation transcriptionnelle de *ptsH* (codant Hpr) semblerait indiquer que si Hpr a un rôle dans l'ATR, il ne serait pas médié par CcpA et la répression catabolique car alors, l'expression de *ptsH* devrait augmenter.

3.4.3. ATR et métabolisme des acides gras

Des changements de la composition membranaire en acides gras durant l'ATR ont été décrits chez quelques bactéries lactiques (*S. sanguis*, *S. gordonii*, *S. salivarius* et *L. casei*, Fozo *et al.*, 2004a ; *S. mutans*, Quivey *et al.*, 2000). Chez les bactéries, la modification de la longueur et du degré de désaturation des acides gras membranaires dépend des enzymes de synthèse (Raetz, 1978). Nos résultats montrent que deux groupes de fonctions impliquées dans le métabolisme des lipides sont affectés par l'ATR, d'une part les gènes *accC* et *fabI* impliqués dans la voie de biosynthèse des acides gras de type II sont induits et d'autre part, les protéines MvaC et MvaS, impliquées dans la synthèse du mévalonate diminuent (figure 27).

L'analyse du contexte génétique suggère que les gènes *accC* et *fabI* font partie d'un opéron de 12 gènes (figure 26), tous impliqués dans la biosynthèse des acides gras. L'induction transcriptionnelle de ces deux gènes suggère donc que l'ensemble de cette voie de biosynthèse est induit en condition acide ce qui devrait conduire à une synthèse d'acides gras plus

importante. Rappelons ici que chez *S. mutans* qui possède une désaturase, la voie de biosynthèse des acides gras de type II est probablement nécessaire au développement de l'ATR puisque la présence de céruléine, inhibiteur de la protéine FabF (figure 27), sensibilise la souche au stress acide (Fozo *et al.*, 2004b).

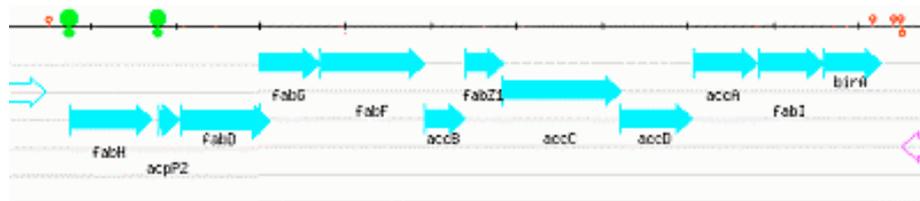


Figure 26 : Contexte génétique des gènes impliqués dans la voie de biosynthèse des acides gras de type II.

Bien qu'ATCC 11842 possède a priori tous les gènes nécessaires à la biosynthèse des acides gras, une interrogation sur la fonctionnalité de cette voie subsiste. En effet, des expériences complémentaires montrent que comme la plupart des souches de *L. bulgaricus*, ATCC 11842 ne se développe pas sans une source externe d'acides gras (le MRS contient de l'acide oléique C_{18:1} ; Partanen *et al.*, 2001). En MRS, la caractérisation des acides gras membranaires (réalisée par G. Lamberet et A. Fernandez, com. pers.) ne révèle aucun changement de la composition en acides gras de la membrane après adaptation acide. Notons cependant qu'en MRS la membrane est composée en majorité d'acide oléique, acide gras largement majoritaire dans le milieu et présent sous forme de Tween 80. Cette source d'acides gras pourrait masquer ou rendre inutile des modifications d'acides gras membranaires, peut-être essentielle dans d'autres milieux. Dans le lait, l'acide oléique ne représente que 26% des acides gras (source : www.fao.org), il serait intéressant de voir si en condition d'ATR en lait, on observe un changement dans la composition membranaire.

Comme mentionné ci-dessus, l'autre groupe de gènes, réprimé en condition d'adaptation, est impliqué dans la voie du mévalonate, voie complète chez ATCC 11842. Cette voie mène à la synthèse d'isoprénoides, impliqués dans la synthèse du peptidoglycane et des acides téchoïques.

Les voies du mévalonate et de synthèse d'acides gras de type II bifurquent au niveau de l'acetyl-CoA (figure 27). Chez *L. bulgaricus*, l'enzyme responsable de la formation d'acetyl-coA à partir du pyruvate (pyruvate dehydrogenase) est absente. La seule voie pour synthétiser ce composé passe par la formation d'acétyl phosphate à partir du pyruvate par la pyruvate oxydase (Pox). Cette enzyme présente deux isoformes dont les quantités varient durant l'ATR : une forme augmente (2,5 fois), l'autre forme diminue (1,9 fois). Mentionnons que la pyruvate oxydase n'est pas connue pour être phosphorylée. Nous avons également montré qu'au niveau transcriptionnel,

pox est induit.. L'induction de *pox* permettrait d'augmenter la quantité d'acetyl-coA produite, et la répression de la voie du mévalonate permettrait ainsi d'orienter la majeure partie de l'acetyl-coA dans la voie de biosynthèse d'acides gras de type II, dont nous proposons qu'elle soit, tout comme chez *S. mutans*, importante pour le développement de l'ATR.

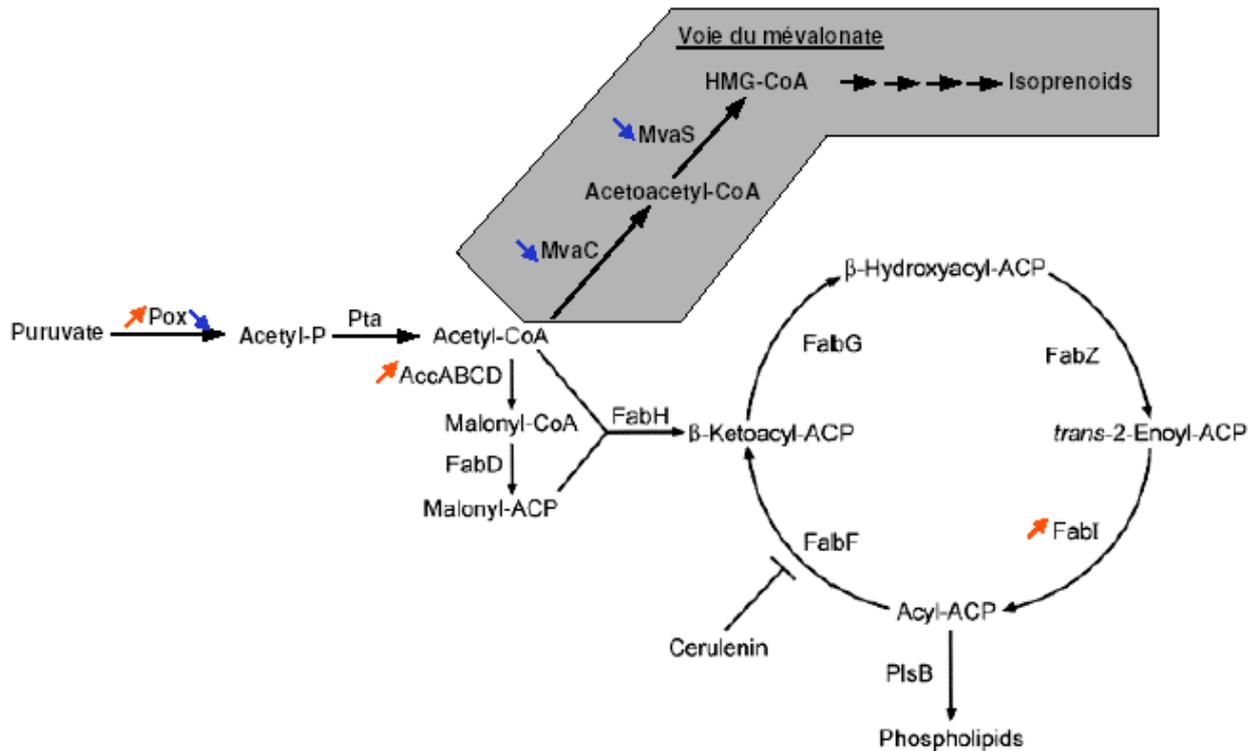


Figure 27 : Voies de biosynthèse des phospholipides et des isoprénoides affectées par l'ATR

↗ Protéine dont l'abondance augmente durant l'ATR. ↘ Protéine dont l'abondance diminue durant l'ATR.

En gris, la voie du mévalonate qui semble défavorisée durant l'ATR. L'effet inhibiteur de la céruléine sur FabF est schématisé.

Adapté de Heath *et al.*, 2001.

Discussion générale

Perspectives

1. Adaptation et dérive génétique

1.1. Evolution du génome

Une des caractéristiques les plus marquantes du génome de *L. bulgaricus* est le nombre de gènes inactivés (12% des gènes du génome, transposases exclues). La présence importante de pseudogènes avait précédemment été décrite, chez les bactéries, pour des pathogènes intracellulaires comme *Rickettsia prowazekii* (>20 % de pseudogènes) ou *Mycobacterium leprae* (>40%) (Andersson et Andersson, 2001 ; Cole *et al.*, 2001). Parmi les bactéries n'étant pas inféodées à un hôte, seule *S. thermophilus*, l'autre bactérie utilisée pour la production de yaourt, présente un nombre élevé de pseudogènes (11% des gènes) (Bolotin *et al.*, 2004). Chez les bactéries indépendantes d'un hôte, le pourcentage de pseudogènes ne dépasse jamais les 5% (Homma *et al.*, 2002). Chez *E. coli*, par exemple, le pourcentage de pseudogènes de la souche O157:H7 est de 3,8 % (Lerat et Ochman, 2004) et chez les lactobacilles, *L. salivarius* ne présente que 3,6 % de gènes inactivés (Claesson *et al.*, 2006). Ce pourcentage est comparable à celui trouvé chez la levure *Saccharomyces cerevisiae*, qui est de 4,8 % (Lafontaine *et al.*, 2004).

Parmi les pathogènes dépendants d'un hôte, Ochman et Davalos (2006) distinguent deux catégories : les pathogènes adaptées depuis longtemps à une vie intracellulaire ou symbiotique et les pathogènes récemment spécialisés à un hôte. Le génome des pathogènes de la première catégorie est extrêmement réduit et les pseudogènes sont rares alors que les représentants de la deuxième catégorie présentent des génomes de taille intermédiaire (2 à 5 Mb) et un grand nombre de pseudogènes (*M. leprae* par exemple) (figure 28). *L. bulgaricus*, par ses caractéristiques génomiques (taille du génome, nombre de pseudogènes, nombre d'IS) se rapproche plus des espèces appartenant à cette deuxième catégorie.

Dans ce deuxième groupe, on retrouve des pathogènes intracellulaires stricts, mais dont la spécialisation à une niche écologique restreinte est relativement récente comme *Shigella flexneri* (Wei *et al.*, 2003), *M. leprae* (Cole *et al.*, 2001) ou *R. prowazekii* (Andersson et Andersson, 2001) et des pathogènes extracellulaires de spectre d'hôte très restreint comme *Bordetella pertussis*, pathogène spécifique de l'homme (Parkhill *et al.*, 2003). Les bactéries appartenant à ce groupe seraient à un stade d'évolution transitoire vers une réduction drastique de la taille de leur génome (Ochman et Davalos, 2006) retrouvé chez les pathogènes de la catégorie 1. A titre d'exemples, la catégorie 1 comprend des pathogènes intracellulaires stricts et des symbiotes

comme *M. genitalium*, parasite de l'homme (580 kb, Fraser *et al.*, 1995), *Phytoplasma asteris*, pathogène intracellulaire de plante (860 kb, Oshima *et al.*, 2004) ou *Blochmannia floridanus*, un endosymbiote d'insecte (706 kb, Gil *et al.*, 2003).

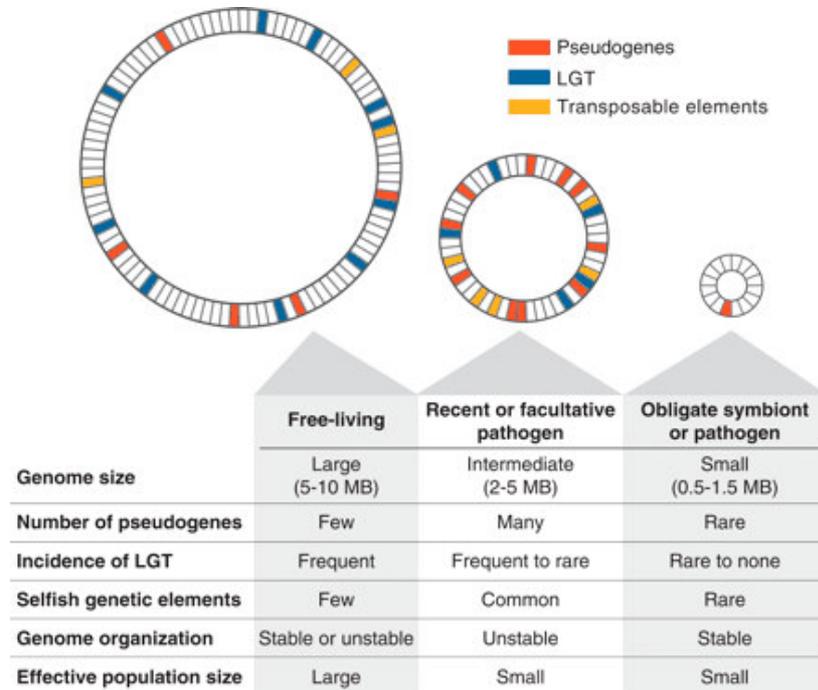


Figure 28 : Caractéristiques des génomes bactériens selon le mode de vie (Ochman et Davalos, 2006).

Différents mécanismes sont impliqués dans l'évolution des génomes des pathogènes dépendants d'un hôte pour leur développement.

Une caractéristique des génomes des bactéries dans les étapes précoces de réduction génomique, partagée par *L. bulgaricus*, est la présence d'un nombre important d'IS (Moran et Plague, 2004). L'hypothèse actuelle est que dans les génomes, le nombre d'IS résulte de deux forces opposées. La première est la tendance naturelle des IS à se multiplier dans le génome par transposition. La deuxième est la pression de sélection contre l'inactivation des gènes bénéfiques pour la souche qui contre-sélectionne les événements de transposition (limitant ainsi de nombres d'IS dans le génome) et favorise les mécanismes d'inactivation des transposases. Lorsqu'une souche évolue d'un mode de vie libre vers une pathogénie obligatoire, la population diminue drastiquement (en général, l'infection d'un hôte ne nécessite qu'une faible quantité de bactéries). La compétition pour la niche écologique est alors beaucoup moins forte, ce qui va relâcher la contre-sélection d'événements de transposition, et par là même favorise la multiplication des IS dans le génome. De plus, vivre dans un environnement riche en nutriments (fournis par l'hôte)

permet à la souche de perdre certains gènes devenus superflus dans ces conditions (Moran et Plague, 2004).

La prolifération des IS a plusieurs conséquences. Les événements de transposition génèrent des séquences répétées qui peuvent être à la base de réarrangements génomiques. Ainsi, chez *S. flexneri* et *B. pertussis*, la plupart des réarrangements observés par comparaison avec des bactéries proches sont bordés par des IS (Wei *et al.*, 2003 ; Parkhill *et al.*, 2003). De plus, chaque événement de transposition est susceptible d'inactiver un ou plusieurs gènes (dans le cas de mutations polaires). Ces gènes, une fois inactivés, seront éliminés, ce qui est probablement une des raisons de la diminution des génomes de pathogènes.

Avec *S. thermophilus*, *L. bulgaricus* est une des seules bactéries non pathogènes à partager les caractéristiques génomiques des pathogènes inféodés à un hôte. *L. bulgaricus*, par ces aspects, pourrait être considéré comme un « parasite » du lait. En effet, les mécanismes potentiellement responsables des modifications génomiques associées à une spécialisation à une niche environnementale se retrouvent dans la sélection par l'homme des souches de *L. bulgaricus* pour la fermentation du lait. Le lait est un milieu riche qui fournit un nombre important de nutriments (acides gras, acides aminés ou peptides, hydrates de carbone...), la pression de sélection est donc moins forte sur certaines voies de biosynthèse. Certains pseudogènes suggèrent que *L. bulgaricus* est issu d'un habitat végétal, et aurait pu contaminer le lait à partir de cet habitat. L'ensemencement par l'homme du lait par une petite quantité de yaourt ou de ferments préparés à partir de milieux riches a reproduit les effets décrits chez les pathogènes : une diminution drastique de la population et une baisse de la pression de sélection pour les gènes nécessaires au mode de vie antérieur (dans ce cas-là, le milieu végétal). Plus que de l'adaptation au milieu, la perte de gènes semble être une dérive génétique liée à la perte de pression de sélection. De plus, les souches ont été sélectionnées au cours du temps pour certaines caractéristiques technologiques (résistance aux phages, acidification rapide, production d'exopolysaccharides), introduisant ainsi un nouveau biais dans la pression de sélection, et accélérant éventuellement la formation de pseudogènes.

Par rapport au schéma général décrit chez les pathogènes, le cas de *L. bulgaricus* présente cependant quelques différences notables : (i) ATCC 11842 présente un nombre important d'IS (3% des gènes) mais 66% d'entre elles sont inactivées. (ii) aucune famille d'IS n'est significativement sur-représentée (la famille ISL7 la plus représentée ne représente que 29% du total) comme nous pourrions l'attendre dans le cas d'une prolifération par événements successifs de transposition et comme décrit chez *Bordetella* par exemple (Parkhill *et al.*, 2003). (iii) aucun

gène n'est inactivé par insertion d'une IS (à part des transposases), ce qui suggère que les IS n'aient pas joué un rôle prépondérant dans l'apparition de pseudogènes. La comparaison avec un génome d'un *L. delbrueckii* d'une autre sous-espèce qui n'aurait pas subi de restriction de niche écologique (*L. delbrueckii lactis* par exemple dont les niches écologiques sont aussi diverses que les animaux ou les végétaux) pourrait déterminer à quel point les phénomènes décrits ci-dessus (apparition de pseudogènes, diminution de la taille du génome, présence importante d'IS) ont influencé la structure du génome de *L. bulgaricus* et si des réarrangements génomiques ont eu lieu et sont liés ou non à la présence d'IS.

1.2. Pseudogènes et adaptation

Outre les IS, une autre source de gènes mutés ou inactivés, découlant du même facteur (i.e. la diminution de la pression de sélection sur certaines fonctions) est l'apparition de mutations ponctuelles (insertion, délétion ou substitution provoquant un codon stop précoce ou un changement du cadre de lecture).

L'apparition de pseudogènes entraîne souvent une perte de fonction et est intuitivement associée à un effet négatif pour les capacités de la souche. Cependant, il a été montré que, dans certains cas, l'apparition de pseudogènes était un atout pour la pathogénie des souches et avait donc un effet positif (Kresse *et al.*, 2003 ; Kallastu *et al.*, 1998 ; Safi *et al.*, 2004). Chez *L. bulgaricus*, un exemple de mutation bénéfique est l'inactivation du régulateur LacR ou de son opérateur, mise en évidence dans plusieurs souches (Germond *et al.*, 2003) et qui se retrouve dans le génome d'ATCC 11842. Cette inactivation permet l'expression constitutive de l'opéron *lac* et ainsi une utilisation optimale du lactose, principale source de carbone du lait. Il est donc difficile de déterminer si l'accumulation de pseudogènes a un effet positif, négatif ou neutre sur les performances de la souche lorsque celle-ci se développe dans le lait.

Une autre question qui se pose vis à vis des pseudogènes est leur inactivation fonctionnelle effective. Le nombre important de gènes inactivés mais encore potentiellement transcrits et traduits pour produire des protéines tronquées et/ou inactives semble être une dépense d'énergie inutile. Mira et Pushker (2005) ont montré que les promoteurs des pseudogènes et notamment les séquences Shine-Dalgarno sont particulièrement dégradées devant les pseudogènes de plusieurs espèces (par exemple, *M. leprae*). Cependant, cette observation n'est pas généralisable à toutes les espèces et ce n'est pas le cas pour *Shigella flexneri* par exemple. L'hypothèse avancée est que les espèces pour lesquelles les séquences Shine-Dalgarno n'ont pas disparu sont celles dont la

spécialisation et par conséquent l'apparition des pseudogènes est la plus récente. La réduction des génomes et l'élimination des pseudogènes pourrait être un moyen de diminuer le coût engendré par la réplication d'ADN non codant donc inutile et entraînerait ainsi la diminution de la taille des génomes.

Notons cependant que l'identification des pseudogènes par analyse bioinformatique peut conduire à des erreurs, comme dans le cas de changement de cadre de lecture programmé (Engelberg-Kulka et Schoulaker-Schwarz, 1994). Un gène fonctionnel sera alors considéré comme pseudogène. A titre d'exemple, il a été montré l'inactivation d'un régulateur fragmenté considéré comme pseudogène chez *S. thermophilus* conduit à un phénotype de résistance au stress oxydant (Fernandez *et al.*, 2004).

L'inactivation différentielle des gènes entre les souches peut être une source de variation phénotypique. Ce phénomène pourrait notamment être à l'origine des différences de tolérance et d'adaptation au stress acide observées entre différentes souches de *L. bulgaricus* (Guillouard *et al.*, 2004). Une autre source de variation de phénotypes entre les souches est l'acquisition de fonctions par transfert horizontal. Au laboratoire, une étude est en cours sur l'analyse de la diversité génomique de souches de *L. bulgaricus*. La méthode utilisée permet de mettre en évidence l'acquisition de gènes par transfert horizontaux mais pas les inactivations différentielles de gènes (C. Grimaldi, comm. pers.). Prochainement, une autre séquence génomique de *L. bulgaricus* sera entièrement disponible, ce qui permettra de comparer les pseudogènes présents dans les deux souches et d'avoir une idée de la variabilité possible entre les contenus génétiques présents dans les différentes souches de *L. bulgaricus*.

2. Réponse adaptative à l'acidité de L. bulgaricus

Les résultats obtenus lors des études sur l'adaptation de *L. bulgaricus* au stress acide, de par les techniques utilisées, sont complémentaires. En effet, par analyse protéomique 2D, les transporteurs et les régulateurs ne sont que très rarement détectés (pI, PM, quantités incompatibles avec la détection ou protéines non solubles). A titre d'illustration, seuls huit régulateurs et un transporteur d'ions (3 sous-unités de l'H⁺-ATPase) sont détectés sur la carte protéomique de référence d'ATCC 11842 (carte réalisée à partir de cultures en phase exponentielle en MPL comprenant 300 protéines caractérisées ; S. Boudebouze, comm. pers.). L'utilisation, pour des raisons techniques, de milieux de culture différents (MPL pour l'étude des

régulateurs et des transporteurs d'ions et MRS pour la 2DE) rend difficile la comparaison de nos résultats obtenus par les deux études. Chez *L. lactis*, les analyses protéomiques réalisées sur l'ATR en milieu riche (M17) et en milieu chimiquement défini (SA) montrent un faible recouvrement entre les protéines induites dans ces deux milieux (Budin-Verneuil *et al.*, 2005b). Il serait donc intéressant d'avoir les données complètes en transcriptomique et en protéomique en MPL ou en MRS pour l'ATR de *L. bulgaricus*. Cependant, l'utilisation de milieux de culture différents permet l'identification des mécanismes pouvant intervenir dans l'ATR dans les deux milieux et ainsi avoir une idée plus exhaustive de ces mécanismes.

2.1. Régulation de l'ATR

Lors de l'étude des systèmes de régulation, nous avons mis en évidence 13 régulateurs dont l'expression est modifiée en condition d'ATR. Il est possible que, par cette méthode, nous n'ayons pas identifié tous les régulateurs impliqués dans l'ATR car le niveau d'expression d'un régulateur n'est pas forcément corrélé avec son activité. Leur activation peut aussi dépendre d'un co-effecteur ou d'une phosphorylation (cas des systèmes à deux composants). Dans ce cas, un régulateur peut être impliqué dans une réponse à un stimulus environnemental sans que son expression varie.

La question qui se pose est comment déterminer les régulons mis en œuvre lors de l'ATR. La disponibilité de la séquence génomique de *L. bulgaricus* ouvre la voie de la transcriptomique par la méthode des puces à ADN. Ces études permettraient d'avoir une vue globale des gènes dont l'expression est modifiée en condition d'ATR et de rechercher une régulation commune à différents gènes par comparaison de leurs séquences promotrices (recherche de motifs conservés).

Il semble aujourd'hui relativement clair que chez les BL, aucun analogue à σ^B de *B. subtilis* ne régule l'expression des gènes impliqués dans l'ATR. L'hypothèse d'une coordination par l'acidité de petits régulons indépendants est renforcée par les résultats obtenus dans les deux études : aucune régulation commune n'a pu être mise en évidence, ni par les études transcriptionnelles ni par les études protéomiques et les fonctions identifiées (métabolisme des acides gras, transports d'ions ou de composés toxiques, réponse au choc thermique) sont peu susceptibles d'être régulés par le même système de régulation. Dans cette étude, deux régulons répondant à l'ATR ont été mis en évidence : le régulon CopY et le régulon HrcA. Les régulateurs impliqués dans la régulation des autres gènes dont l'expression varie en condition

d'ATR restent à déterminer ainsi que l'étendue et l'importance pour la mise en place de l'ATR des régulons contrôlés par les régulateurs dont l'expression est affectée par l'ATR.

De part le faible nombre de ses régulateurs et de ses systèmes de résistance, *L. bulgaricus* semblait un bon modèle d'étude pour l'ATR. Au vu des résultats obtenus, la régulation de la réponse adaptative de *L. bulgaricus* apparaît plus complexe qu'attendue. Il semble que cette régulation fasse intervenir non seulement un certain nombre de régulateurs transcriptionnels, mais aussi des phénomènes d'atténuation de la transcription et des régulations post-traductionnelles, visualisées en 2DE.

2.2. ATR et régulon du choc thermique

L'induction des protéines de réponse au choc thermique en condition d'ATR a été largement documentée chez les BL (Hartke *et al.*, 1997 ; Lemos *et al.*, 2001 ; Len *et al.*, 2004 ; Lim *et al.*, 2000). Cette induction est cohérente avec l'observation que l'ATR rend *L. bulgaricus* plus résistant à un choc thermique (E.M. Lim, comm. pers.). Cependant, les résultats obtenus lors de l'étude de l'expression des régulateurs ne semble pas, au premier abord, cohérente avec ceux obtenus lors de la validation transcriptionnelle des résultats protéomiques. En effet, le gène *hrcA*, codant le répresseur des gènes de la réponse au choc thermique et situé en opéron avec *dnaK*, n'est pas induit en MPL après 90 min à pH 5,25 alors que *dnaK* est induit en MRS après 40 min à pH 3,9. Une hypothèse pourrait être que le régulon HrcA n'est pas induit lors de l'ATR en MPL. Cependant, une étude en 2DE sur l'adaptation de la souche Vi1010 en MPL montre que GroES, GroEL et DnaK sont induites par l'ATR (30 min d'adaptation Lim *et al.*, 2000). Cette apparente contradiction pourrait être due à un effet de cinétique. En effet, dans plusieurs études, il a été montré que l'induction des gènes de la réponse au choc thermique lors de l'ATR est transitoire (Lemos *et al.*, 2001 ; E.M. Lim, comm. pers.). Les résultats obtenus en MRS montrent qu'à 40 min, le facteur d'induction des protéines chaperons a déjà diminué d'un facteur ~ 3. Il est possible qu'en MPL, après 90 min d'adaptation, la quantité de transcrits de l'opéron *hrcA-dnaK* soit revenue à son état initial (c'est-à-dire au niveau de l'échantillon non adapté) après une induction transitoire.

La question qui se pose face à l'ensemble de ces protéines ou gènes induits au cours de l'ATR est de déterminer s'ils contribuent de façon directe et significative à la réponse adaptative. La réponse à cette question peut sembler a priori simple, une solution étant d'inactiver un gène et

de tester l'effet sur la tolérance à l'acidité par rapport à la souche sauvage avec et sans adaptation. De telles études ont déjà été réalisées à partir de données protéomiques (A. Budin-Verneuil chez *L. lactis*, comm. pers. ; T. Sasaki chez *L. gasseri*, comm. pers.) ou à partir de gènes candidats identifiés par analyse *in silico* (Azcarate-Peril *et al.*, 2004). La situation la plus fréquente est d'observer une sensibilité au stress acide plus importante dans le mutant (avec des degrés de sensibilité plus ou moins marqués) que pour la souche sauvage en absence d'adaptation. Cependant, le plus souvent, l'adaptation à l'acidité supprime tout ou partie de la sensibilité du mutant par rapport à la souche sauvage. Ces observations renforcent l'hypothèse que l'ATR pourrait être multifactorielle et mettrait en jeu des mécanismes en partie redondants ou pouvant compenser la perte de l'un d'entre eux. Pour mieux comprendre l'importance respective des mécanismes de la mise en place de l'ATR, il serait peut-être nécessaire d'étudier des mutants affectés dans plusieurs gènes ou fonctions. Cette approche pourrait être facilitée par la connaissance des systèmes de régulation de l'ATR. Une possibilité serait d'étudier le phénotype de mutants de régulation, l'inactivation d'un régulateur permettant alors d'affecter l'expression de l'ensemble de son régulon.

Les travaux réalisés chez *L. bulgaricus* ATCC 11842 ont permis d'identifier un certain nombre de gènes ou protéines candidats potentiellement impliqués dans l'ATR. Cette bactérie, de part la réduction drastique de ses capacités codantes liée à l'inactivation de nombreux gènes, pourrait être un bon modèle d'étude de l'ATR. En effet, une hypothèse est que *L. bulgaricus* aurait conservé les mécanismes essentiels pour l'ATR, les fonctions redondantes ou superflues étant éliminées. Ne resteraient alors que les mécanismes dont la perte aurait un effet néfaste pour la bactérie et celle-ci serait alors contre-sélectionnée.

Références bibliographiques

- Abe, K., Hayashi, H. & Maloney, P. C. (1996). Exchange of aspartate and alanine. Mechanism for development of a proton-motive force in bacteria. *J Biol Chem* 271, 3079-3084.
- Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M. & Aksoy, S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* 32, 402-407.
- Alakomi, H. L., Skytta, E., Saarela, M., Mattila-Sandholm, T., Latva-Kala, K. & Helander, I. M. (2000). Lactic acid permeabilizes gram-negative bacteria by disrupting the outer membrane. *Appl Environ Microbiol* 66, 2001-2005.
- Alam, M., Midtvedt, T. & Uribe, A. (1994). Differential cell kinetics in the ileum and colon of germfree rats. *Scand J Gastroenterol* 29, 445-451.
- Allardet-Servent, A., Michaux-Charachon, S., Jumas-Bilak, E., Karayan, L. & Ramuz, M. (1993). Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. *J Bacteriol* 175, 7869-7874.
- Al-Mushrif, S. & Jones, B. M. (1998). A study of the prevalence of hydrogen peroxide generating Lactobacilli in bacterial vaginosis: the determination of H₂O₂ concentrations generated, in vitro, by isolated strains and the levels found in vaginal secretions of women with and without infection. *J Obstet Gynaecol* 18, 63-67.
- Altermann, E., Russell, W. M., Azcarate-Peril, M. A. & other authors (2005). Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc Natl Acad Sci U S A* 102, 3906-3912.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402.
- Andersson, J. O. & Andersson, S. G. (2001). Pseudogenes, junk DNA, and the dynamics of *Rickettsia* genomes. *Mol Biol Evol* 18, 829-839.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews* 29, 231-262.
- Au, N., Kuester-Schoeck, E., Mandava, V. & other authors (2005). Genetic composition of the *Bacillus subtilis* SOS system. *J Bacteriol* 187, 7655-7666.
- Azcarate-Peril, M. A., Altermann, E., Hoover-Fitzula, R. L., Cano, R. J. & Klaenhammer, T. R. (2004). Identification and inactivation of genetic loci involved with *Lactobacillus acidophilus* acid tolerance. *Appl Environ Microbiol* 70, 5315-5322.
- Azcarate-Peril, M. A., McAuliffe, O., Altermann, E., Lick, S., Russell, W. M. & Klaenhammer, T. R. (2005). Microarray analysis of a two-component regulatory system involved in acid resistance and proteolytic activity in *Lactobacillus acidophilus*. *Appl Environ Microbiol* 71, 5794-5804.
- Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915-1920.
- Bailey, T. L. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2, 28-36.

- Bakker, E. P. & Harold, F. M. (1980). Energy coupling to potassium transport in *Streptococcus faecalis*. Interplay of ATP and the protonmotive force. *J Biol Chem* 255, 433-440.
- Barriere, C., Veiga-da-Cunha, M., Pons, N., Guedon, E., van Hijum, S. A., Kok, J., Kuipers, O. P., Ehrlich, D. S. & Renault, P. (2005). Fructose utilization in *Lactococcus lactis* as a model for low-GC gram-positive bacteria: its regulator, signal, and DNA-binding site. *J Bacteriol* 187, 3752-3761.
- Bateman, A., Birney, E., Cerruti, L. & other authors (2002). The Pfam protein families database. *Nucleic Acids Res* 30, 276-280.
- Benachour, A., Muller, C., Dabrowski-Coton, M., Le Breton, Y., Giard, J. C., Rince, A., Auffray, Y. & Hartke, A. (2005). The *Enterococcus faecalis* sigV protein is an extracytoplasmic function sigma factor contributing to survival following heat, acid, and ethanol treatments. *J Bacteriol* 187, 1022-1035.
- Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340, 783-795.
- Benson, A. K. & Haldenwang, W. G. (1993). The sigma B-dependent promoter of the *Bacillus subtilis* sigB operon is induced by heat shock. *J Bacteriol* 175, 1929-1935.
- Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M. & other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141-147.
- Bentley, S. D. & Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annu Rev Genet* 38, 771-792.
- Berg, R. D. (1996). The indigenous gastrointestinal microflora. *Trends Microbiol* 4, 430-435.
- Bibel, D. J. (1988). Elie Metchnikoff's bacillus of long life. *ASM News* 54, 661-665.
- Bijlsma, J. J. & Groisman, E. A. (2003). Making informed decisions: regulatory interactions between two-component systems. *Trends Microbiol* 11, 359-366.
- Bischoff, M., Dunman, P., Kormanec, J., Macapagal, D., Murphy, E., Mounts, W., Berger-Bachi, B. & Projan, S. (2004). Microarray-based analysis of the *Staphylococcus aureus* sigmaB regulon. *J Bacteriol* 186, 4085-4099.
- Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A. & other authors (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453-1474.
- Boekhorst, J., Siezen, R. J., Zwahlen, M. C. & other authors (2004). The complete genomes of *Lactobacillus plantarum* and *Lactobacillus johnsonii* reveal extensive differences in chromosome organization and gene content. *Microbiology* 150, 3601-3611.
- Bolotin, A., Quinquis, B., Renault, P. & other authors (2004). Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* 22, 1554-1558.
- Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S. D. & Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* 11, 731-753.
- Borodina, I. & Nielsen, J. (2005). From genomes to in silico cells via metabolic networks. *Curr Opin Biotechnol* 16, 350-355.

- Boyd, D. A., Cvitkovitch, D. G., Bleiweis, A. S., Kiriukhin, M. Y., Debabov, D. V., Neuhaus, F. C. & Hamilton, I. R. (2000a). Defects in D-alanyl-lipoteichoic acid synthesis in *Streptococcus mutans* results in acid sensitivity. *J Bacteriol* 182, 6055-6065.
- Boyd, D. A., Peters, G. A., Ng, L. & Mulvey, M. R. (2000b). Partial characterization of a genomic island associated with the multidrug resistance region of *Salmonella enterica* Typhimurium DT104. *FEMS Microbiol Lett* 189, 285-291.
- Boylan, S. A., Redfield, A. R., Brody, M. S. & Price, C. W. (1993). Stress-induced activation of the sigma B transcription factor of *Bacillus subtilis*. *J Bacteriol* 175, 7931-7937.
- Brigulla, M., Hoffmann, T., Krisp, A., Volker, A., Bremer, E. & Volker, U. (2003). Chill induction of the SigB-dependent general stress response in *Bacillus subtilis* and its contribution to low-temperature adaptation. *J Bacteriol* 185, 4305-4314.
- Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. (2003). The MerR family of transcriptional regulators. *FEMS Microbiol Rev* 27, 145-163.
- Buck, B. L., Altermann, E., Svingerud, T. & Klaenhammer, T. R. (2005). Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 71, 8344-8351.
- Budin-Verneuil, A., Maguin, E., Auffray, Y., Ehrlich, S. D. & Pichereau, V. (2004). An essential role for arginine catabolism in the acid tolerance of *Lactococcus lactis* MG1363. *Le Lait* 84, 61-68.
- Budin-Verneuil, A., Maguin, E., Auffray, Y., Ehrlich, S. D. & Pichereau, V. (2005a). Transcriptional analysis of the cyclopropane fatty acid synthase gene of *Lactococcus lactis* MG1363 at low pH. *FEMS Microbiol Lett* 250, 189-194.
- Budin-Verneuil, A., Pichereau, V., Auffray, Y., Ehrlich, S. D. & Maguin, E. (2005b). Proteomic characterization of the acid tolerance response in *Lactococcus lactis* MG1363. *Proteomics* 5, 4794-4807.
- Busch, W. & Saier, M. H., Jr. (2002). The transporter classification (TC) system, 2002. *Crit Rev Biochem Mol Biol* 37, 287-337.
- Campos, E., Baldoma, L., Aguilar, J. & Badia, J. (2004). Regulation of expression of the divergent *ulaG* and *ulaABCDEF* operons involved in LaAscorbate dissimilation in *Escherichia coli*. *J Bacteriol* 186, 1720-1728.
- Cannon, J. P., Lee, T. A., Bolanos, J. T. & Danziger, L. H. (2005). Pathogenic relevance of *Lactobacillus*: a retrospective review of over 200 cases. *Eur J Clin Microbiol Infect Dis* 24, 31-40.
- Carbone, A., Zinovyev, A. & Kepes, F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics* 19, 2005-2015.
- Chaillou, S., Champomier-Verges, M. C., Cornet, M. & other authors (2005). The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. *Nat Biotechnol*.
- Champomier Verges, M. C., Zuniga, M., Morel-Deville, F., Perez-Martinez, G., Zagorec, M. & Ehrlich, S. D. (1999). Relationships between arginine degradation, pH and survival in *Lactobacillus sakei*. *FEMS Microbiol Lett* 180, 297-304.
- Champomier-Verges, M. C., Chaillou, S., Cornet, M. & Zagorec, M. (2001). *Lactobacillus sakei*: recent developments and future prospects. *Res Microbiol* 152, 839-848.

- Chang, Y. Y. & Cronan, J. E., Jr. (1999). Membrane cyclopropane fatty acid content is a major factor in acid resistance of *Escherichia coli*. *Mol Microbiol* 33, 249-259.
- Chastanet, A., Derre, I., Nair, S. & Msadek, T. (2004). *clpB*, a novel member of the *Listeria monocytogenes* CtsR regulon, is involved in virulence but not in general stress tolerance. *J Bacteriol* 186, 1165-1174.
- Chastanet, A., Fert, J. & Msadek, T. (2003a). Comparative genomics reveal novel heat shock regulatory mechanisms in *Staphylococcus aureus* and other Gram-positive bacteria. *Mol Microbiol* 47, 1061-1073.
- Chastanet, A. & Msadek, T. (2003b). ClpP of *Streptococcus salivarius* is a novel member of the dually regulated class of stress response genes in gram-positive bacteria. *J Bacteriol* 185, 683-687.
- Chastanet, A., Prudhomme, M., Claverys, J. P. & Msadek, T. (2001). Regulation of *Streptococcus pneumoniae* *clp* genes and their role in competence development and stress survival. *J Bacteriol* 183, 7295-7307.
- Chattopadhyay, M. K., Tabor, C. W. & Tabor, H. (2003). Polyamines protect *Escherichia coli* cells from the toxic effect of oxygen. *Proc Natl Acad Sci U S A* 100, 2261-2265.
- Chen, Y. Y., Weaver, C. A. & Burne, R. A. (2000). Dual functions of *Streptococcus salivarius* urease. *J Bacteriol* 182, 4667-4669.
- Cheo, D. L., Bayles, K. W. & Yasbin, R. E. (1991). Cloning and characterization of DNA damage-inducible promoter regions from *Bacillus subtilis*. *J Bacteriol* 173, 1696-1703.
- Chervaux, C., Ehrlich, S. D. & Maguin, E. (2000). Physiological study of *Lactobacillus delbrueckii* subsp. *bulgaricus* strains in a novel chemically defined medium. *Appl Environ Microbiol* 66, 5306-5311.
- Claesson, M. J., Li, Y., Leahy, S. & other authors (2006). Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc Natl Acad Sci U S A*.
- Cole, S. T., Eiglmeier, K., Parkhill, J. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007-1011.
- Collins, M. D., Rodrigues, U., Ash, C., Aguirre, M., Farrow, J. A. E., Martinez-Murcia, A., Phillips, B. A., Williams, A. M. & Wallbanks, S. (1991). Phylogenetic analysis of the genus *Lactobacillus* and related lactic acid bacteria as determined by reverse transcriptase sequencing of 16S rRNA. *FEMS Microbiology Letters* 77, 5-12.
- Corcoran, B. M., Stanton, C., Fitzgerald, G. F. & Ross, R. P. (2005). Survival of probiotic lactobacilli in acidic environments is enhanced in the presence of metabolizable sugars. *Appl Environ Microbiol* 71, 3060-3067.
- Corpet, F., Gouzy, J. & Kahn, D. (1998). The ProDom database of protein domain families. *Nucleic Acids Res* 26, 323-326.
- Cotter, P. D. & Hill, C. (2003). Surviving the acid test: responses of gram-positive bacteria to low pH. *Microbiol Mol Biol Rev* 67, 429-453, table of contents.
- Coucheney, F., Gal, L., Beney, L., Lherminier, J., Gervais, P. & Guzzo, J. (2005). A small HSP, Lo18, interacts with the cell membrane and modulates lipid physical state under heat shock conditions in a lactic acid bacterium. *Biochim Biophys Acta* 1720, 92-98.

- Courtin, P., Monnet, V. & Rul, F. (2002). Cell-wall proteinases PrtS and PrtB have a different role in *Streptococcus thermophilus*/*Lactobacillus bulgaricus* mixed cultures in milk. *Microbiology* 148, 3413-3421.
- Craig, E. A., Gambill, B. D. & Nelson, R. J. (1993). Heat-Shock Proteins - Molecular Chaperones of Protein Biogenesis. *Microbiological Reviews* 57, 402-414.
- Crowley, P. J., Svensater, G., Snoep, J. L., Bleiweis, A. S. & Brady, L. J. (2004). An *ffh* mutant of *Streptococcus mutans* is viable and able to physiologically adapt to low pH in continuous culture. *FEMS Microbiol Lett* 234, 315-324.
- Cummings, J. H. & Macfarlane, G. T. (1997). Colonic microflora: nutrition and health. *Nutrition* 13, 476-478.
- Cunin, R., Glansdorff, N., Pierard, A. & Stalon, V. (1986). Biosynthesis and metabolism of arginine in bacteria. *Microbiol Rev* 50, 314-352.
- Dalton, T. L. & Scott, J. R. (2004). *CovS* inactivates *CovR* and is required for growth under conditions of general stress in *Streptococcus pyogenes*. *J Bacteriol* 186, 3928-3937.
- Darmon, E., Noone, D., Masson, A., Bron, S., Kuipers, O. P., Devine, K. M. & van Dijk, J. M. (2002). A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated *CssRS* two-component system of *Bacillus subtilis*. *J Bacteriol* 184, 5661-5671.
- Dashper, S. G. & Reynolds, E. C. (1992). pH regulation by *Streptococcus mutans*. *J Dent Res* 71, 1159-1165.
- d'Aubenton Carafa, Y., Brody, E. & Thermes, C. (1990). Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J Mol Biol* 216, 835-858.
- Daubin, V., Lerat, E. & Perriere, G. (2003). The source of laterally transferred genes in bacterial genomes. *Genome Biol* 4, R57.
- De Angelis, M., Bini, L., Pallini, V., Cocconcelli, P. S. & Gobbetti, M. (2001). The acid-stress response in *Lactobacillus sanfranciscensis* CB1. *Microbiology* 147, 1863-1873.
- De Angelis, M. & Gobbetti, M. (2004). Environmental stress responses in *Lactobacillus*: a review. *Proteomics* 4, 106-122.
- De Angelis, M., Mariotti, L., Rossi, J., Servili, M., Fox, P. F., Rollan, G. & Gobbetti, M. (2002). Arginine catabolism by sourdough lactic acid bacteria: purification and characterization of the arginine deiminase pathway enzymes from *Lactobacillus sanfranciscensis* CB1. *Appl Environ Microbiol* 68, 6193-6201.
- de Roissart, H. & Luquet, F. M. (1994). *Bactéries Lactiques, Aspects fondamentaux et technologiques*. Uriage: Lorica.
- de Vos, W. M., Bron, P. A. & Kleerebezem, M. (2004). Post-genomics of lactic acid bacteria and other food-grade bacteria to discover gut functionality. *Curr Opin Biotechnol* 15, 86-93.
- del Campo, R., Bravo, D., Canton, R., Ruiz-Garbajosa, P., Garcia-Albiach, R., Montesi-Libois, A., Yuste, F. J., Abaira, V. & Baquero, F. (2005). Scarce evidence of yogurt lactic acid bacteria in human feces after daily yogurt consumption by healthy volunteers. *Appl Environ Microbiol* 71, 547-549.

- Delmas, F., Pierre, F., Coucheney, F., Divies, C. & Guzzo, J. (2001). Biochemical and physiological studies of the small heat shock protein Lo18 from the lactic acid bacterium *Oenococcus oeni*. *J Mol Microbiol Biotechnol* 3, 601-610.
- Derre, I., Rapoport, G., Devine, K., Rose, M. & Msadek, T. (1999a). ClpE, a novel type of HSP100 ATPase, is part of the CtsR heat shock regulon of *Bacillus subtilis*. *Mol Microbiol* 32, 581-593.
- Derre, I., Rapoport, G. & Msadek, T. (1999b). CtsR, a novel regulator of stress and heat shock response, controls *clp* and molecular chaperone gene expression in gram-positive bacteria. *Mol Microbiol* 31, 117-131.
- Derre, I., Rapoport, G. & Msadek, T. (2000). The CtsR regulator of stress response is active as a dimer and specifically degraded in vivo at 37 degrees C. *Mol Microbiol* 38, 335-347.
- Derzelle, S., Hallet, B., Francis, K. P., Ferain, T., Delcour, J. & Hols, P. (2000). Changes in *cspL*, *cspP*, and *cspC* mRNA abundance as a function of cold shock and growth phase in *Lactobacillus plantarum*. *J Bacteriol* 182, 5105-5113.
- Deuerling, E., Paeslack, B. & Schumann, W. (1995). The *ftsH* gene of *Bacillus subtilis* is transiently induced after osmotic and temperature upshift. *J Bacteriol* 177, 4105-4112.
- Diaz-Torres, M. L. & Russell, R. R. B. (2001). HtrA protease and processing of extracellular proteins of *Streptococcus mutans*. *Fems Microbiology Letters* 204, 23-28.
- Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2, 414-424.
- Drider, D., Bekal, S. & Prevost, H. (2004). Genetic organization and expression of citrate permease in lactic acid bacteria. *Genet Mol Res* 3, 273-281.
- Duwat, P., Sourice, S., Cesselin, B. & other authors (2001). Respiration capacity of the fermenting bacterium *Lactococcus lactis* and its positive effects on growth and survival. *J Bacteriol* 183, 4509-4516.
- Elgrably-Weiss, M., Schlosser-Silverman, E., Rosenshine, I. & Altuvia, S. (2006). DeoT, a DeoR-type transcriptional regulator of multiple target genes. *FEMS Microbiol Lett* 254, 141-148.
- Engelberg-Kulka, H. & Schoulaker-Schwarz, R. (1994). Regulatory implications of translational frameshifting in cellular gene expression. *Mol Microbiol* 11, 3-8.
- Enright, M. C. & Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol* 7, 482-487.
- Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K. & Bairoch, A. (2002). The PROSITE database, its status in 2002. *Nucleic Acids Res* 30, 235-238.
- Faustoferri, R. C., Hahn, K., Weiss, K. & Quivey, R. G., Jr. (2005). Smx nuclease is the major, low-pH-inducible apurinic/aprimidinic endonuclease in *Streptococcus mutans*. *J Bacteriol* 187, 2705-2714.
- Fayol-Messaoudi, D., Berger, C. N., Coconnier-Polter, M. H., Lievin-Le Moal, V. & Servin, A. L. (2005). pH-, Lactic acid-, and non-lactic acid-dependent activities of probiotic *Lactobacilli* against *Salmonella enterica* Serovar Typhimurium. *Appl Environ Microbiol* 71, 6008-6013.
- Ferdows, M. S. & Barbour, A. G. (1989). Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. *Proc Natl Acad Sci U S A* 86, 5969-5973.

- Fernandez, A., Thibessard, A., Borges, F., Gintz, B., Decaris, B. & Leblond-Bourget, N. (2004). Characterization of oxidative stress-resistant mutants of *Streptococcus thermophilus* CNRZ368. *Arch Microbiol* 182, 364-372.
- Fleischmann, R. D., Adams, M. D., White, O. & other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496-512.
- Forsyth, R. A., Haselbeck, R. J., Ohlsen, K. L. & other authors (2002). A genome-wide strategy for the identification of essential genes in *Staphylococcus aureus*. *Mol Microbiol* 43, 1387-1400.
- Fox, P. F. (1993). Cheese : an overview. In *Cheese : chemistry, physics and microbiology*, pp. 1-36. Edited by P. F. Fox. London: Chapman and Hall.
- Fozo, E. M., Kajfasz, J. K. & Quivey, R. G., Jr. (2004a). Low pH-induced membrane fatty acid alterations in oral bacteria. *FEMS Microbiol Lett* 238, 291-295.
- Fozo, E. M. & Quivey, R. G., Jr. (2004b). Shifts in the membrane fatty acid profile of *Streptococcus mutans* enhance survival in acidic environments. *Appl Environ Microbiol* 70, 929-936.
- Fozo, E. M., Quivey, R. G., Jr., Faustoferri, R., Monahan, K. & Marquis, R. (2004c). The *fabM* gene product of *Streptococcus mutans* is responsible for the synthesis of monounsaturated fatty acids and is necessary for survival at low pH
Shifts in membrane fatty acid profiles associated with acid adaptation of *Streptococcus mutans*. *J Bacteriol* 186, 4152-4158.
- Fraser, C. M., Gocayne, J. D., White, O. & other authors (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397-403.
- Frees, D. & Ingmer, H. (1999). ClpP participates in the degradation of misfolded protein in *Lactococcus lactis*. *Molecular Microbiology* 31, 79-87.
- Frees, D., Thomsen, L. E. & Ingmer, H. (2005). *Staphylococcus aureus* ClpYQ plays a minor role in stress survival. *Arch Microbiol* 183, 286-291.
- Frees, D., Vogensen, F. K. & Ingmer, H. (2003). Identification of proteins induced at low pH in *Lactococcus lactis*. *Int J Food Microbiol* 87, 293-300.
- Gale, E. F. (1940). Enzymes Concerned in the Primary Utilization of Amino Acids by Bacteria. *Bacteriol Rev* 4, 135-176.
- Garcia-Quintans, N., Magni, C., de Mendoza, D. & Lopez, P. (1998). The citrate transport system of *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis* is induced by acid stress. *Appl Environ Microbiol* 64, 850-857.
- Gardy, J. L., Laird, M. R., Chen, F., Rey, S., Walsh, C. J., Ester, M. & Brinkman, F. S. (2005). PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* 21, 617-623.
- Gerlt, J. A. & Babbitt, P. C. (2000). Can sequence determine function? *Genome Biol* 1, REVIEWS0005.
- Germond, J.-E., Lapierre, L., Delley, M., Mollet, B., Felis, G. E. & Dellaglio, F. (2003). Evolution of the Bacterial Species *Lactobacillus delbrueckii*: A Partial Genomic Study with Reflections on Prokaryotic Species Concept
10.1093/molbev/msg012. *Mol Biol Evol* 20, 93-104.

- Ghosal, D., Omelchenko, M. V., Gaidamakova, E. K. & other authors (2005). How radiation kills cells: survival of *Deinococcus radiodurans* and *Shewanella oneidensis* under oxidative stress. *FEMS Microbiol Rev* 29, 361-375.
- Gil, R., Silva, F. J., Pereto, J. & Moya, A. (2004). Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* 68, 518-537.
- Gil, R., Silva, F. J., Zientz, E. & other authors (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc Natl Acad Sci U S A* 100, 9388-9393.
- Glaser, P., Frangeul, L., Buchrieser, C. & other authors (2001). Comparative genomics of *Listeria* species. *Science* 294, 849-852.
- Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., Hutchison, C. A., 3rd, Smith, H. O. & Venter, J. C. (2006). Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A* 103, 425-430.
- Gottesman, S., Wickner, S. & Maurizi, M. R. (1997). Protein quality control: Triage by chaperones and proteases. *Genes & Development* 11, 815-823.
- Gouesbet, G., Jan, G. & Boyaval, P. (2002). Two-dimensional electrophoresis study of *Lactobacillus delbrueckii* subsp. *bulgaricus* thermotolerance. *Appl Environ Microbiol* 68, 1055-1063.
- Graham, M. R., Smoot, L. M., Migliaccio, C. A. & other authors (2002). Virulence control in group A *Streptococcus* by a two-component gene regulatory system: global expression profiling and in vivo infection modeling. *Proc Natl Acad Sci U S A* 99, 13855-13860.
- Granato, D., Perotti, F., Masserey, I., Rouvet, M., Golliard, M., Servin, A. & Brassart, D. (1999). Cell surface-associated lipoteichoic acid acts as an adhesion factor for attachment of *Lactobacillus johnsonii* La1 to human enterocyte-like Caco-2 cells. *Appl Environ Microbiol* 65, 1071-1077.
- Grandvalet, C., Coucheney, F., Beltramo, C. & Guzzo, J. (2005). CtsR is the master regulator of stress response gene expression in *Oenococcus oeni*. *J Bacteriol* 187, 5614-5623.
- Grangette, C., Nutten, S., Palumbo, E. & other authors (2005). Enhanced antiinflammatory capacity of a *Lactobacillus plantarum* mutant synthesizing modified teichoic acids. *Proc Natl Acad Sci U S A* 102, 10321-10326.
- Grobben, G. J., Chin-Joe, I., Kitzen, V. A., Boels, I. C., Boer, F., Sikkema, J., Smith, M. R. & de Bont, J. A. (1998). Enhancement of Exopolysaccharide Production by *Lactobacillus delbrueckii* subsp. *bulgaricus* NCFB 2772 with a Simplified Defined Medium. *Appl Environ Microbiol* 64, 1333-1337.
- Grundy, F. J. & Henkin, T. M. (2003). The T box and S box transcription termination control systems. *Front Biosci* 8, d20-31.
- Guarner, F. & Malagelada, J. R. (2003). Gut flora in health and disease. *Lancet* 361, 512-519.
- Guarner, F., Perdigon, G., Corthier, G., Salminen, S., Koletzko, B. & Morelli, L. (2005). Should yoghurt cultures be considered probiotic? *Br J Nutr* 93, 783-786.
- Guillouard, I., Lim, E. M., Van de Guchte, M., Grimaldi, C., Penaud, S. & Maguin, E. (2004). Tolerance and adaptive acid stress response of *Lactobacillus delbrueckii* ssp. *bulgaricus*. *Le Lait* 84, 1-6.
- Gupta, R. S. & Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theor Popul Biol* 61, 423-434.

- Gutierrez, J. A., Crowley, P. J., Cvitkovitch, D. G., Brady, L. J., Hamilton, I. R., Hillman, J. D. & Bleiweis, A. S. (1999). Streptococcus mutans ffh, a gene encoding a homologue of the 54 kDa subunit of the signal recognition particle, is involved in resistance to acid stress. *Microbiology* 145 (Pt 2), 357-366.
- Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R. & Goebel, W. (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal Escherichia coli isolates. *Microb Pathog* 8, 213-225.
- Haller, D., Serrant, P., Granato, D., Schiffrin, E. J. & Blum, S. (2002). Activation of human NK cells by staphylococci and lactobacilli requires cell contact-dependent costimulation by autologous monocytes. *Clin Diagn Lab Immunol* 9, 649-657.
- Hamilton-Miller, J. M. (2003). The role of probiotics in the treatment and prevention of Helicobacter pylori infection. *Int J Antimicrob Agents* 22, 360-366.
- Hanke, M. T. & Koessler, K. K. (1924). STUDIES ON PROTEINOGENOUS AMINES. XVII. ON THE FACULTY OF NORMAL INTESTINAL BACTERIA TO FORM TOXIC AMINES. *J Biol Chem* 59, 835-853.
- Hanna, M. N., Ferguson, R. J., Li, Y. H. & Cvitkovitch, D. G. (2001). uvrA is an acid-inducible gene involved in the adaptive response to low pH in Streptococcus mutans. *Journal of Bacteriology* 183, 5964-5973.
- Harold, F. M., Pavlasova, E. & Baarda, J. R. (1970). A transmembrane pH gradient in Streptococcus faecalis: origin, and dissipation by proton conductors and N,N'-dicyclohexylcarbodiimide. *Biochim Biophys Acta* 196, 235-244.
- Hartig, E., Geng, H., Hartmann, A., Hubacek, A., Munch, R., Ye, R. W., Jahn, D. & Nakano, M. M. (2004). Bacillus subtilis ResD induces expression of the potential regulatory genes yclJK upon oxygen limitation. *J Bacteriol* 186, 6477-6484.
- Hartke, A., Bouche, S., Laplace, J.-M., Benachour, A., Boutibonnes, P. & Auffray, Y. (1995). UV-inducible proteins and UV-induced cross-protection against acid, ethanol, H₂O₂ or heat treatments in Lactococcus lactis subsp. lactis. *Archives of Microbiology* 163, 329-336.
- Hartke, A., Frere, J., Boutibonnes, P. & Auffray, Y. (1997). Differential induction of the chaperonin GroEL and the Co-chaperonin GroES by heat, acid, and UV-irradiation in Lactococcus lactis subsp. lactis. *Curr Microbiol* 34, 23-26.
- Heath, R. J., White, S. W. & Rock, C. O. (2001). Lipid biosynthesis as a target for antibacterial agents. *Prog Lipid Res* 40, 467-497.
- Helmann, J. D. (2002). The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 46, 47-110.
- Herrgard, M. J., Covert, M. W. & Palsson, B. O. (2004). Reconstruction of microbial transcriptional regulatory networks. *Current Opinion in Biotechnology* 15, 70-77.
- Higuchi, T., Hayashi, H. & Abe, K. (1997). Exchange of glutamate and gamma-aminobutyrate in a Lactobacillus strain. *J Bacteriol* 179, 3362-3364.
- Hill, M. J. (1997). Intestinal flora and endogenous vitamin synthesis. *Eur J Cancer Prev* 6 Suppl 1, S43-45.

- Hoebeke, M., Nicolas, P. & Bessieres, P. (2003). MuGeN: simultaneous exploration of multiple genomes and computer analysis results. *Bioinformatics* 19, 859-864.
- Hoesl, C. E. & Altwein, J. E. (2005). The probiotic approach: an alternative treatment option in urology. *Eur Urol* 47, 288-296.
- Holtmann, G., Bakker, E. P., Uozumi, N. & Bremer, E. (2003). KtrAB and KtrCD: two K⁺ uptake systems in *Bacillus subtilis* and their role in adaptation to hypertonicity. *J Bacteriol* 185, 1289-1298.
- Holzapfel, W. H., Haberer, P., Geisen, R., Bjorkroth, J. & Schillinger, U. (2001). Taxonomy and important features of probiotic microorganisms in food and nutrition. *Am J Clin Nutr* 73, 365S-373S.
- Homma, K., Fukuchi, S., Kawabata, T., Ota, M. & Nishikawa, K. (2002). A systematic investigation identifies a significant number of probable pseudogenes in the *Escherichia coli* genome. *Gene* 294, 25-33.
- Homuth, G., Masuda, S., Mogk, A., Kobayashi, Y. & Schumann, W. (1997). The *dnaK* operon of *Bacillus subtilis* is heptacistronic. *J Bacteriol* 179, 1153-1164.
- Hoper, D., Volker, U. & Hecker, M. (2005). Comprehensive characterization of the contribution of individual SigB-dependent general stress genes to stress resistance of *Bacillus subtilis*. *J Bacteriol* 187, 2810-2826.
- Houry, W. A. (2001). Chaperone-assisted protein folding in the cell cytoplasm. *Curr Protein Pept Sci* 2, 227-244.
- Howell, A., Dubrac, S., Andersen, K. K., Noone, D., Fert, J., Msadek, T. & Devine, K. (2003). Genes controlled by the essential YycG/YycF two-component system of *Bacillus subtilis* revealed through a novel hybrid regulator approach. *Mol Microbiol* 49, 1639-1655.
- Huang, X., Decatur, A., Sorokin, A. & Helmann, J. D. (1997). The *Bacillus subtilis* sigma(X) protein is an extracytoplasmic function sigma factor contributing to survival at high temperature. *J Bacteriol* 179, 2915-2921.
- Hughes, D. (2000). Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biol* 1, REVIEWS0006.
- Hutson, J. Y. & Downing, M. (1968). Pyrimidine biosynthesis in *Lactobacillus leichmannii*. *J Bacteriol* 96, 1249-1254.
- Hyyrylainen, H. L., Bolhuis, A., Darmon, E. & other authors (2001). A novel two-component regulatory system in *Bacillus subtilis* for the survival of severe secretion stress. *Mol Microbiol* 41, 1159-1172.
- Ishii, N., Robert, M., Nakayama, Y., Kanai, A. & Tomita, M. (2004). Toward large-scale modeling of the microbial cell for computer simulation. *J Biotechnol* 113, 281-294.
- Isolauri, E., Salminen, S. & Ouwehand, A. C. (2004). Microbial-gut interactions in health and disease. *Probiotics. Best Pract Res Clin Gastroenterol* 18, 299-313.
- Jiang, S. M., Cieslewicz, M. J., Kasper, D. L. & Wessels, M. R. (2005). Regulation of virulence by a two-component system in group B streptococcus. *J Bacteriol* 187, 1105-1113.
- Johnson, D. C., Dean, D. R., Smith, A. D. & Johnson, M. K. (2005). Structure, function, and formation of biological iron-sulfur clusters. *Annu Rev Biochem* 74, 247-281.

- Jordan, I. K., Kondrashov, F. A., Rogozin, I. B., Tatusov, R. L., Wolf, Y. I. & Koonin, E. V. (2001). Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal and eukaryotic proteins. *Genome Biol* 2, RESEARCH0053.
- Kakinuma, Y. (1998). Inorganic cation transport and energy transduction in *Enterococcus hirae* and other streptococci. *Microbiol Mol Biol Rev* 62, 1021-1045.
- Kallastu, A., Horak, R. & Kivisaar, M. (1998). Identification and characterization of IS1411, a new insertion sequence which causes transcriptional activation of the phenol degradation genes in *Pseudomonas putida*. *J Bacteriol* 180, 5306-5312.
- Kappes, R. M., Kempf, B., Kneip, S., Boch, J., Gade, J., Meier-Wagner, J. & Bremer, E. (1999). Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. *Mol Microbiol* 32, 203-216.
- Karp, P. D., Paley, S. & Romero, P. (2002). The Pathway Tools software. *Bioinformatics* 18 Suppl 1, S225-232.
- Kashket, E. R. & Barker, S. L. (1977). Effects of potassium ions on the electrical and pH gradients across the membrane of *Streptococcus lactis* cells. *J Bacteriol* 130, 1017-1023.
- Kashket, E. R., Blanchard, A. G. & Metzger, W. C. (1980). Proton motive force during growth of *Streptococcus lactis* cells. *J Bacteriol* 143, 128-134.
- Kazmierczak, M. J., Mithoe, S. C., Boor, K. J. & Wiedmann, M. (2003). *Listeria monocytogenes* sigma B regulates stress response and virulence functions. *J Bacteriol* 185, 5722-5734.
- Kempf, B. & Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch Microbiol* 170, 319-330.
- Kieser, H. M., Kieser, T. & Hopwood, D. A. (1992). A combined genetic and physical map of the *Streptomyces coelicolor* A3(2) chromosome. *J Bacteriol* 174, 5496-5507.
- Kilstrup, M. & Martinussen, J. (1998). A transcriptional activator, homologous to the *Bacillus subtilis* PurR repressor, is required for expression of purine biosynthetic genes in *Lactococcus lactis*. *J Bacteriol* 180, 3907-3916.
- Kirstein, J., Zuhlke, D., Gerth, U., Turgay, K. & Hecker, M. (2005). A tyrosine kinase and its activator control the activity of the CtsR heat shock repressor in *B. subtilis*. *Embo J* 24, 3435-3445.
- Klaenhammer, T. R. (1988). Bacteriocins of lactic acid bacteria. *Biochimie* 70, 337-349.
- Klaenhammer, T. R. (1993). Genetics of bacteriocins produced by lactic acid bacteria. *FEMS Microbiol Rev* 12, 39-85.
- Klaenhammer, T. R., Barrangou, R., Buck, B. L., Azcarate-Peril, M. A. & Altermann, E. (2005). Genomic features of lactic acid bacteria effecting bioprocessing and health. *FEMS Microbiol Rev* 29, 393-409.
- Kleerebezem, M., Boekhorst, J., van Kranenburg, R. & other authors (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A* 100, 1990-1995.
- Kobayashi, H. (1985). A proton-translocating ATPase regulates pH of the bacterial cytoplasm. *J Biol Chem* 260, 72-76.

- Kobayashi, H., Murakami, N. & Unemoto, T. (1982). Regulation of the cytoplasmic pH in *Streptococcus faecalis*. *J Biol Chem* 257, 13246-13252.
- Kobayashi, H., Suzuki, T. & Unemoto, T. (1986). Streptococcal cytoplasmic pH is regulated by changes in amount and activity of a proton-translocating ATPase. *J Biol Chem* 261, 627-630.
- Kobayashi, K., Ehrlich, S. D., Albertini, A. & other authors (2003). Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci U S A* 100, 4678-4683.
- Koch, A. L. (2003a). Were Gram-positive rods the first bacteria? *Trends Microbiol* 11, 166-170.
- Koch, H. G., Moser, M. & Muller, M. (2003b). Signal recognition particle-dependent protein targeting, universal to all kingdoms of life. *Rev Physiol Biochem Pharmacol* 146, 55-94.
- Koebmann, B. J., Nilsson, D., Kuipers, O. P. & Jensen, P. R. (2000). The membrane-bound H(+)-ATPase complex is essential for growth of *Lactococcus lactis*. *J Bacteriol* 182, 4738-4743.
- Koonin, E. V. (2003). Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* 1, 127-136.
- Korber, P., Stahl, J. M., Nierhaus, K. H. & Bardwell, J. C. (2000). Hsp15: a ribosome-associated heat shock protein. *Embo J* 19, 741-748.
- Korner, H., Sofia, H. J. & Zumft, W. G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiol Rev* 27, 559-592.
- Koski, L. B., Morton, R. A. & Golding, G. B. (2001). Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol Biol Evol* 18, 404-412.
- Kovacs, T., Hargitai, A., Kovacs, K. L. & Mecs, I. (1998). pH-dependent activation of the alternative transcriptional factor sigmaB in *Bacillus subtilis*. *FEMS Microbiol Lett* 165, 323-328.
- Kresse, A. U., Dinesh, S. D., Larbig, K. & Romling, U. (2003). Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Mol Microbiol* 47, 145-158.
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305, 567-580.
- Kruger, E. & Hecker, M. (1998). The first gene of the *Bacillus subtilis* clpC operon, ctsR, encodes a negative regulator of its own operon and other class III heat shock genes. *J Bacteriol* 180, 6681-6688.
- Kunst, F., Ogasawara, N., Moszer, I. & other authors (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390, 249-256.
- Lafay, B., Atherton, J. C. & Sharp, P. M. (2000). Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori*. *Microbiology* 146 (Pt 4), 851-860.
- Lafontaine, I., Fischer, G., Talla, E. & Dujon, B. (2004). Gene relics in the genome of the yeast *Saccharomyces cerevisiae*. *Gene* 335, 1-17.

- Lamothe, G. T., Jolly, L., Mollet, B. & Stinglele, F. (2002). Genetic and biochemical characterization of exopolysaccharide biosynthesis by *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Arch Microbiol* 178, 218-228.
- Laplace, J. M., Sauvageot, N., Hartke, A. & Auffray, Y. (1999). Characterization of *Lactobacillus collinoides* response to heat, acid and ethanol treatments. *Applied Microbiology and Biotechnology* 51, 659-663.
- Lawrence, J. G. & Hartl, D. L. (1992). Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. *Genetics* 131, 753-760.
- Lawrence, J. G. & Ochman, H. (2002). Reconciling the many faces of lateral gene transfer. *Trends in Microbiology* 10, 1-4.
- Lehto, E. M. & Salminen, S. J. (1997). Inhibition of *Salmonella typhimurium* adhesion to Caco-2 cell cultures by *Lactobacillus* strain GG spent culture supernate: only a pH effect? *FEMS Immunol Med Microbiol* 18, 125-132.
- Lemos, J. A., Chen, Y. Y. & Burne, R. A. (2001). Genetic and physiologic analysis of the *groE* operon and role of the *HrcA* repressor in stress gene regulation and acid tolerance in *Streptococcus mutans*. *J Bacteriol* 183, 6074-6084.
- Lemos, J. A. C. & Burne, R. A. (2002). Regulation and physiological significance of ClpC and ClpP in *Streptococcus mutans*. *Journal of Bacteriology* 184, 6357-6366.
- Len, A. C. L., Harty, D. W. S. & Jacques, N. A. (2004). Stress-responsive proteins are upregulated in *Streptococcus mutans* during acid tolerance. *Microbiology-Sgm* 150, 1339-1351.
- Lerat, E. & Ochman, H. (2004). Psi-Phi: exploring the outer limits of bacterial pseudogenes. *Genome Res* 14, 2273-2278.
- Lick, S., Drescher, K. & Heller, K. J. (2001). Survival of *Lactobacillus delbrueckii* subsp. *bulgaricus* and *Streptococcus thermophilus* in the terminal ileum of fistulated Gottingen minipigs. *Appl Environ Microbiol* 67, 4137-4143.
- Lim, E. M., Ehrlich, S. D. & Maguin, E. (2000). Identification of stress-inducible proteins in *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Electrophoresis* 21, 2557-2561.
- Lindahl, T. & Andersson, A. (1972). Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry* 11, 3618-3623.
- Lister, J. (1873). A further contribution to the natural history of bacteria and the germ theory of fermentative changes. *Quarterly Microbiological Sciences* 13, 380-408.
- Liu, M., van Enkevort, F. H. & Siezen, R. J. (2005). Genome update: lactic acid bacteria genome sequencing is booming. *Microbiology* 151, 3811-3814.
- Liu, S., Pritchard, G., Hardman, M. & Pilone, G. (1995). Occurrence of Arginine Deiminase Pathway Enzymes in Arginine Catabolism by Wine Lactic Acid Bacteria. *Appl Environ Microbiol* 61, 310-316.
- Liu, S.-Q. & Pilone, G. J. (1998). A REVIEW : Arginine metabolism in wine lactic acid bacteria and its practical significance. *Journal of Applied Microbiology* 84, 315-327.
- Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻($\Delta\Delta C(T)$) Method. *Methods* 25, 402-408.

- Lorca, G. L., Raya, R. R., Taranto, M. P. & de Valdez, G. F. (1998). Adaptive acid tolerance response in *Lactobacillus acidophilus*. *Biotechnology Letters* 20, 239-241.
- Lowe, T. M. & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25, 955-964.
- Lucas, P., Landete, J., Coton, M., Coton, E. & Lonvaud-Funel, A. (2003). The tyrosine decarboxylase operon of *Lactobacillus brevis* IOEB 9809: characterization and conservation in tyramine-producing bacteria. *FEMS Microbiol Lett* 229, 65-71.
- Lyon, G. J. & Novick, R. P. (2004). Peptide signaling in *Staphylococcus aureus* and other Gram-positive bacteria. *Peptides* 25, 1389-1403.
- Lysenko, E., Ogura, T. & Cutting, S. M. (1997). Characterization of the *ftsH* gene of *Bacillus subtilis*. *Microbiology* 143 (Pt 3), 971-978.
- Magalhaes, P. P., Paulino, T. P., Thedei, G., Jr. & Ciancaglini, P. (2005). Kinetic characterization of P-type membrane ATPase from *Streptococcus mutans*. *Comp Biochem Physiol B Biochem Mol Biol* 140, 589-597.
- Magalhaes, P. P., Paulino, T. P., Thedei, G., Jr., Larson, R. E. & Ciancaglini, P. (2003). A 100 kDa vanadate and lansoprazole-sensitive ATPase from *Streptococcus mutans* membrane. *Arch Oral Biol* 48, 815-824.
- Makarova, K. S., Mironov, A. A. & Gelfand, M. S. (2001). Conservation of the binding site for the arginine repressor in all bacterial lineages. *Genome Biol* 2, RESEARCH0013.
- Manca de Nadra, M. C., Pesce de Ruiz Holgado, A. A. & Oliver, G. (1988). Arginine dihydrolase pathway in *Lactobacillus buchneri*: a review. *Biochimie* 70, 367-374.
- Marchler-Bauer, A. & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* 32, W327-331.
- Marrakchi, H., Zhang, Y. M. & Rock, C. O. (2002). Mechanistic diversity and regulation of Type II fatty acid synthesis. *Biochem Soc Trans* 30, 1050-1055.
- Martin, M. G., Magni, C., de Mendoza, D. & Lopez, P. (2005). CitI, a transcription factor involved in regulation of citrate metabolism in lactic acid bacteria. *J Bacteriol* 187, 5146-5155.
- Martin, M. G., Sender, P. D., Peiru, S., de Mendoza, D. & Magni, C. (2004). Acid-inducible transcription of the operon encoding the citrate lyase complex of *Lactococcus lactis* Biovar diacetylactis CRL264. *J Bacteriol* 186, 5649-5660.
- Martin-Galiano, A. J., Overweg, K., Ferrandiz, M. J., Reuter, M., Wells, J. M. & de la Campa, A. G. (2005). Transcriptional analysis of the acid tolerance response in *Streptococcus pneumoniae*. *Microbiology* 151, 3935-3946.
- Marty-Teyssset, C., de la Torre, F. & Garel, J. (2000). Increased production of hydrogen peroxide by *Lactobacillus delbrueckii* subsp. *bulgaricus* upon aeration: involvement of an NADH oxidase in oxidative stress. *Appl Environ Microbiol* 66, 262-267.
- Marty-Teyssset, C., Posthuma, C., Lolkema, J. S., Schmitt, P., Divies, C. & Konings, W. N. (1996). Proton motive force generation by citrolactic fermentation in *Leuconostoc mesenteroides*. *J Bacteriol* 178, 2178-2185.

- Mater, D. D., Bretigny, L., Firmesse, O., Flores, M. J., Mogenet, A., Bresson, J. L. & Corthier, G. (2005). *Streptococcus thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus* survive gastrointestinal transit of healthy volunteers consuming yogurt. *FEMS Microbiol Lett* 250, 185-187.
- McCracken, V. J. & Lorenz, R. G. (2001). The gastrointestinal ecosystem: a precarious alliance among epithelium, immunity and microbiota. *Cell Microbiol* 3, 1-11.
- McDonald, L. C., Fleming, H. P. & Hassan, H. M. (1990). Acid Tolerance of *Leuconostoc mesenteroides* and *Lactobacillus plantarum*. *Appl Environ Microbiol* 56, 2120-2124.
- McLean, M. J., Wolfe, K. H. & Devine, K. M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J Mol Evol* 47, 691-696.
- Medigue, C., Rouxel, T., Vigier, P., Henaut, A. & Danchin, A. (1991). Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol* 222, 851-856.
- Merino, E. & Yanofsky, C. (2005). Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet* 21, 260-264.
- Metchnikoff, E. (1907). *The prolongation of life*. London: Heinemann.
- Mills, D. A., Rawsthorne, H., Parker, C., Tamir, D. & Makarova, K. (2005). Genomic analysis of *Oenococcus oeni* PSU-1 and its relevance to winemaking. *FEMS Microbiol Rev* 29, 465-475.
- Mira, A., Ochman, H. & Moran, N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17, 589-596.
- Mira, A. & Pushker, R. (2005). The silencing of pseudogenes. *Mol Biol Evol* 22, 2135-2138.
- Mira De Orduna, R., Patchett, M. L., Liu, S. Q. & Pilone, G. J. (2001). Growth and arginine metabolism of the wine lactic acid bacteria *Lactobacillus buchneri* and *Oenococcus oeni* at different pH values and arginine concentrations. *Appl Environ Microbiol* 67, 1657-1662.
- Miyazawa, E., Iwabuchi, A. & Yoshida, T. (1996). Phytate breakdown and apparent absorption of phosphorus, calcium and magnesium in germfree and conventionalized rats. *Nutrition Research* 16, 603-613.
- Mogk, A., Homuth, G., Scholz, C., Kim, L., Schmid, F. X. & Schumann, W. (1997). The GroE chaperonin machine is a major modulator of the CIRCE heat shock regulon of *Bacillus subtilis*. *Embo J* 16, 4579-4590.
- Molenaar, D., Bosscher, J. S., ten Brink, B., Driessen, A. J. & Konings, W. N. (1993). Generation of a proton motive force by histidine decarboxylation and electrogenic histidine/histamine antiport in *Lactobacillus buchneri*. *J Bacteriol* 175, 2864-2870.
- Molenaar, D., Bringel, F., Schuren, F. H., de Vos, W. M., Siezen, R. J. & Kleerebezem, M. (2005). Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J Bacteriol* 187, 6119-6127.
- Mora, D., Fortina, M. G., Parini, C., Ricci, G., Gatti, M., Giraffa, G. & Manachini, P. L. (2002). Genetic diversity and technological properties of *Streptococcus thermophilus* strains isolated from dairy products. *J Appl Microbiol* 93, 278-287.

- Mora, D., Maguin, E., Masiero, M., Parini, C., Ricci, G., Manachini, P. L. & Daffonchio, D. (2004). Characterization of urease genes cluster of *Streptococcus thermophilus*. *J Appl Microbiol* 96, 209-219.
- Mora, D., Monnet, C., Parini, C., Guglielmetti, S., Mariani, A., Pintus, P., Molinari, F., Daffonchio, D. & Manachini, P. L. (2005). Urease biogenesis in *Streptococcus thermophilus*. *Res Microbiol* 156, 897-903.
- Moran, N. A. & Plague, G. R. (2004). Genomic changes following host restriction in bacteria. *Curr Opin Genet Dev* 14, 627-633.
- Morel-Deville, F., Fauvel, F. & Morel, P. (1998). Two-component signal-transducing systems involved in stress responses and vancomycin susceptibility in *Lactobacillus sakei*. *Microbiology* 144, 2873-2883.
- Morikawa, K., Inose, Y., Okamura, H., Maruyama, A., Hayashi, H., Takeyasu, K. & Ohta, T. (2003). A new staphylococcal sigma factor in the conserved gene cassette: functional significance and implication for the evolutionary processes. *Genes Cells* 8, 699-712.
- Morrison, D. A. & Lee, M. S. (2000). Regulation of competence for genetic transformation in *Streptococcus pneumoniae*: a link between quorum sensing and DNA processing genes. *Res Microbiol* 151, 445-451.
- Moszer, I., Rocha, E. P. & Danchin, A. (1999). Codon usage and lateral gene transfer in *Bacillus subtilis*. *Curr Opin Microbiol* 2, 524-528.
- Mrazek, J. & Karlin, S. (1999). Detecting alien genes in bacterial genomes. *Ann N Y Acad Sci* 870, 314-329.
- Mushegian, A. R. & Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A* 93, 10268-10273.
- Nachin, L., Loiseau, L., Expert, D. & Barras, F. (2003). SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *Embo J* 22, 427-437.
- Nachin, L., Nannmark, U. & Nystrom, T. (2005). Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J Bacteriol* 187, 6265-6272.
- Nagafuchi, S., Takahashi, T., Yajima, T., Kuwata, T., Hirayama, K. & Itoh, K. (1999). Strain dependency of the immunopotentiating activity of *Lactobacillus delbrueckii* subsp. *bulgaricus*. *Biosci Biotechnol Biochem* 63, 474-479.
- Nakano, S., Erwin, K. N., Ralle, M. & Zuber, P. (2005). Redox-sensitive transcriptional control by a thiol/disulphide switch in the global regulator, Spx. *Mol Microbiol* 55, 498-510.
- Nicoloff, H., Arsene-Ploetze, F., Malandain, C., Kleerebezem, M. & Bringel, F. (2004). Two arginine repressors regulate arginine biosynthesis in *Lactobacillus plantarum*. *J Bacteriol* 186, 6059-6069.
- Nilsson, D., Lauridsen, A. A., Tomoyasu, T. & Ogura, T. (1994). A *Lactococcus lactis* gene encodes a membrane protein with putative ATPase activity that is homologous to the essential *Escherichia coli* ftsH gene product. *Microbiology* 140 (Pt 10), 2601-2610.
- Ochman, H. (2002). Bacterial evolution: chromosome arithmetic and geometry. *Curr Biol* 12, R427-428.
- Ochman, H. & Davalos, L. M. (2006). The nature and dynamics of bacterial genomes. *Science* 311, 1730-1733.

- O'Connell-Motherway, M., van Sinderen, D., Morel-Deville, F., Fitzgerald, G. F., Ehrlich, S. D. & Morel, P. (2000). Six putative two-component regulatory systems isolated from *Lactococcus lactis* subsp. *cremoris* MG1363. *Microbiology* 146 (Pt 4), 935-947.
- Ogawa, M., Shimizu, K., Nomoto, K., Tanaka, R., Hamabata, T., Yamasaki, S., Takeda, T. & Takeda, Y. (2001). Inhibition of in vitro growth of Shiga toxin-producing *Escherichia coli* O157:H7 by probiotic *Lactobacillus* strains due to production of lactic acid. *Int J Food Microbiol* 68, 135-140.
- Olsen, E. B., Russell, J. B. & Henick-Kling, T. (1991). Electrogenic L-malate transport by *Lactobacillus plantarum*: a basis for energy derivation from malolactic fermentation. *J Bacteriol* 173, 6199-6206.
- Oommen, B. S., McMahon, D. J., Oberg, C. J., Broadbent, J. R. & Strickland, M. (2002). Proteolytic specificity of *Lactobacillus delbrueckii* subsp. *bulgaricus* influences functional properties of mozzarella cheese. *J Dairy Sci* 85, 2750-2758.
- Oshima, K., Kakizawa, S., Nishigawa, H. & other authors (2004). Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat Genet* 36, 27-29.
- O'Sullivan, E. & Condon, S. (1999). Relationship between acid tolerance, cytoplasmic pH, and ATP and H⁺-ATPase levels in chemostat cultures of *Lactococcus lactis*. *Appl Environ Microbiol* 65, 2287-2293.
- Overbeek, R., Larsen, N., Walunas, T. & other authors (2003). The ERGO genome analysis and discovery system. *Nucleic Acids Res* 31, 164-171.
- Parkhill, J., Sebahia, M., Preston, A. & other authors (2003). Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* 35, 32-40.
- Partanen, L., Marttinen, N. & Alatosava, T. (2001). Fats and fatty acids as growth factors for *Lactobacillus delbrueckii*. *Syst Appl Microbiol* 24, 500-506.
- Pasteur, L. (1857). Mémoire sur la fermentation appelée lactique. *CR Séances Académiques de Sciences* 45, 913-916.
- Paulsen, I. T., Banerjee, L., Myers, G. S. & other authors (2003). Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science* 299, 2071-2074.
- Pedersen, A. G., Jensen, L. J., Brunak, S., Staerfeldt, H. H. & Ussery, D. W. (2000). A DNA structural atlas for *Escherichia coli*. *J Mol Biol* 299, 907-930.
- Perdigon, G., Maldonado Galdeano, C., Valdez, J. C. & Medici, M. (2002). Interaction of lactic acid bacteria with the gut immune system. *Eur J Clin Nutr* 56 Suppl 4, S21-26.
- Petersohn, A., Brigulla, M., Haas, S., Hoheisel, J. D., Volker, U. & Hecker, M. (2001). Global analysis of the general stress response of *Bacillus subtilis*. *J Bacteriol* 183, 5617-5631.
- Petry, S., Furlan, S., Waghorne, E., Saulnier, L., Cerning dagger, J. & Maguin, E. (2003). Comparison of the thickening properties of four *Lactobacillus delbrueckii* subsp. *bulgaricus* strains and physicochemical characterization of their exopolysaccharides. *FEMS Microbiol Lett* 221, 285-291.
- Phadtare, S. (2004). Recent developments in bacterial cold-shock response. *Curr Issues Mol Biol* 6, 125-136.

- Pieterse, B., Leer, R. J., Schuren, F. H. & van der Werf, M. J. (2005). Unravelling the multiple effects of lactic acid stress on *Lactobacillus plantarum* by transcription profiling. *Microbiology* 151, 3881-3894.
- Pilone, G. J. & Kunkee, R. E. (1976). Stimulatory Effect of Malo-Lactic Fermentation on the Growth Rate of *Leuconostoc oenos*. *Appl Environ Microbiol* 32, 405-408.
- Poolman, B., Molenaar, D., Smid, E. J., Ubbink, T., Abee, T., Renault, P. P. & Konings, W. N. (1991). Malolactic fermentation: electrogenic malate uptake and malate/lactate antiport generate metabolic energy. *J Bacteriol* 173, 6030-6037.
- Poquet, I., Saint, V., Seznec, E., Simoes, N., Bolotin, A. & Gruss, A. (2000). HtrA is the unique surface housekeeping protease in *Lactococcus lactis* and is required for natural protein processing. *Molecular Microbiology* 35, 1042-1051.
- Prescott, L. M., Harley, J. P. & Klein, D. A. (1999). *Microbiology*, 4th edn. New York: WCB/McGraw-Hill.
- Presser, K. A., Ratkowsky, D. A. & Ross, T. (1997). Modelling the growth rate of *Escherichia coli* as a function of pH and lactic acid concentration. *Applied and Environmental Microbiology* 63, 2355-2360.
- Pridmore, R. D., Berger, B., Desiere, F. & other authors (2004). The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc Natl Acad Sci U S A* 101, 2512-2517.
- Qi, Y., Kobayashi, Y. & Hulett, F. M. (1997). The *pst* operon of *Bacillus subtilis* has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the *pho* regulon. *J Bacteriol* 179, 2534-2539.
- Quivey, R. G., Jr., Faustoferri, R., Monahan, K. & Marquis, R. (2000). Shifts in membrane fatty acid profiles associated with acid adaptation of *Streptococcus mutans*. *FEMS Microbiol Lett* 189, 89-92.
- Quivey, R. G., Faustoferri, R. C., Clancy, K. A. & Marquis, R. E. (1995). Acid Adaptation in *Streptococcus-Mutans* Ua159 Alleviates Sensitization to Environmental-Stress Due to RecA Deficiency. *Fems Microbiology Letters* 126, 257-261.
- Raetz, C. R. (1978). Enzymology, genetics, and regulation of membrane phospholipid synthesis in *Escherichia coli*. *Microbiol Rev* 42, 614-659.
- Ragan, M. A. (2001). On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol Lett* 201, 187-191.
- Rallu, F., Gruss, A., Ehrlich, S. D. & Maguin, E. (2000). Acid- and multistress-resistant mutants of *Lactococcus lactis* : identification of intracellular stress signals. *Mol Microbiol* 35, 517-528.
- Rammelsberg, M., Müller, E. & Radler, F. (1990). Caseicin 80: purification and characterization of a new bacteriocin from *Lactobacillus casei*. *Archives of Microbiology* 154, 249-252.
- Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R. & Tobes, R. (2005). The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 69, 326-356.
- Rasko, D. A., Myers, G. S. & Ravel, J. (2005). Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics* 6, 2.

- Recsei, P. A. & Snell, E. E. (1972). Histidine decarboxylaseless mutants of *Lactobacillus* 30a: isolation and growth properties. *J Bacteriol* 112, 624-626.
- Reid, G. & Burton, J. (2002). Use of *Lactobacillus* to prevent infection by pathogenic bacteria. *Microbes and Infection* 4, 319-324.
- Reid, G., Jass, J., Sebulsky, M. T. & McCormick, J. K. (2003). Potential uses of probiotics in clinical practice. *Clin Microbiol Rev* 16, 658-672.
- Renault, P., Gaillardin, C. & Heslot, H. (1988). Role of malolactic fermentation in lactic acid bacteria. *Biochimie* 70, 375-379.
- Rezaiki, L., Cesselin, B., Yamamoto, Y., Vido, K., van West, E., Gaudu, P. & Gruss, A. (2004). Respiration metabolism reduces oxidative and acid stress to improve long-term survival of *Lactococcus lactis*. *Mol Microbiol* 53, 1331-1342.
- Richmond, C. S., Glasner, J. D., Mau, R., Jin, H. & Blattner, F. R. (1999). Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res* 27, 3821-3835.
- Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem* 277, 12507-12515.
- Ritz, D., Patel, H., Doan, B., Zheng, M., Aslund, F., Storz, G. & Beckwith, J. (2000). Thioredoxin 2 is involved in the oxidative stress response in *Escherichia coli*. *J Biol Chem* 275, 2505-2512.
- Rojas, M., Ascencio, F. & Conway, P. L. (2002). Purification and characterization of a surface protein from *Lactobacillus fermentum* 104R that binds to porcine small intestinal mucus and gastric mucin. *Appl Environ Microbiol* 68, 2330-2336.
- Rollan, G., Lorca, G. L. & Font de Valdez, G. (2003). Arginine catabolism and acid tolerance response in *Lactobacillus reuteri* isolated from sourdough. *Food Microbiology* 20, 313-319.
- Ross, R. P., Morgan, S. & Hill, C. (2002). Preservation and fermentation: past, present and future. *Int J Food Microbiol* 79, 3-16.
- Russell, J. B. & DiezGonzalez, F. (1998). The effects of fermentation acids on bacterial growth. In *Advances in Microbial Physiology*, Vol 39, pp. 205-234.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944-945.
- Safi, H., Barnes, P. F., Lakey, D. L., Shams, H., Samten, B., Vankayalapati, R. & Howard, S. T. (2004). IS6110 functions as a mobile, monocyte-activated promoter in *Mycobacterium tuberculosis*. *Mol Microbiol* 52, 999-1012.
- Sakoh, M., Ito, K. & Akiyama, Y. (2005). Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from *Escherichia coli*. *J Biol Chem* 280, 33305-33310.
- Salema, M., Poolman, B., Lolkema, J. S., Dias, M. C. & Konings, W. N. (1994). Uniport of monoanionic L-malate in membrane vesicles from *Leuconostoc oenos*. *Eur J Biochem* 225, 289-295.
- Salminen, M. K., Rautelin, H., Tynkkynen, S., Poussa, T., Saxelin, M., Valtonen, V. & Jarvinen, A. (2004). *Lactobacillus* bacteremia, clinical significance, and patient outcome, with special focus on probiotic *L. rhamnosus* GG. *Clin Infect Dis* 38, 62-69.

- Salminen, S. & Arvilommi, H. (2002). Safety of Lactobacillus strains used as probiotic agents. [Reply]. *Clinical Infectious Diseases* 34, 1284-1285.
- Sanders, J. W., Leenhouts, K., Burghoorn, J., Brands, J. R., Venema, G. & Kok, J. (1998). A chloride-inducible acid resistance mechanism in *Lactococcus lactis* and its regulation. *Mol Microbiol* 27, 299-310.
- Sasaki, T., Ito, Y. & Sasaki, Y. (1993). Electrotransformation of *Lactobacillus delbrueckii* subsp. *bulgaricus*. *FEMS Microbiol Rev* 12, P8.
- Savijoki, K., Ingmer, H., Frees, D., Vogensen, F. K., Palva, A. & Varmanen, P. (2003). Heat and DNA damage induction of the LexA-like regulator HdiR from *Lactococcus lactis* is mediated by RecA and ClpP. *Mol Microbiol* 50, 609-621.
- Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annu Rev Microbiol* 47, 597-626.
- Schell, M. A., Karmirantzou, M., Snel, B. & other authors (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci U S A* 99, 14422-14427.
- Schelp, E., Worley, S., Monzingo, A. F., Ernst, S. & Robertus, J. D. (2001). pH-induced structural changes regulate histidine decarboxylase activity in *Lactobacillus* 30a. *J Mol Biol* 306, 727-732.
- Schleifer, K. H., Kraus, J., Dvorak, C., Kilpper-Bälz, R., Collins, M. D. & Fischer, W. (1985). Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus* gen. nov. *Syst Appl Microbiol* 6, 183-195.
- Schleifer, K. H. & Ludwig, W. (1995). Phylogeny of the genus *Lactobacillus* and related genera. *System Appl Microbiol* 18, 461-467.
- Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. & Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* 32, D431-433.
- Schumacher, M. A. & Brennan, R. G. (2002). Structural mechanisms of multidrug recognition and regulation by bacterial multidrug transcription factors. *Mol Microbiol* 45, 885-893.
- Schumann, W. (2003). The *Bacillus subtilis* heat shock stimulon. *Cell Stress Chaperones* 8, 207-217.
- Schuurman-Wolters, G. K. & Poolman, B. (2005). Substrate specificity and ionic regulation of GlnPQ from *Lactococcus lactis*. An ATP-binding cassette transporter with four extracytoplasmic substrate-binding domains. *J Biol Chem* 280, 23785-23790.
- Scott, C., Rawsthorne, H., Upadhyay, M., Shearman, C. A., Gasson, M. J., Guest, J. R. & Green, J. (2000). Zinc uptake, oxidative stress and the FNR-like proteins of *Lactococcus lactis*. *FEMS Microbiol Lett* 192, 85-89.
- Serror, P., Dervyn, R., Ehrlich, S. D. & Maguin, E. (2003). csp-like genes of *Lactobacillus delbrueckii* ssp. *bulgaricus* and their response to cold shock. *FEMS Microbiol Lett* 226, 323-330.
- Serror, P., Sasaki, T., Ehrlich, S. D. & Maguin, E. (2002). Electrotransformation of *Lactobacillus delbrueckii* subsp. *bulgaricus* and *L. delbrueckii* subsp. *lactis* with various plasmids. *Appl Environ Microbiol* 68, 46-52.

- Servin, A. L. (2004). Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev* 28, 405-440.
- Sharp, P. M. & Li, W. H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15, 1281-1295.
- Shelef, L. A. (1994). Antimicrobial Effects of Lactates - a Review. *Journal of Food Protection* 57, 445-450.
- Silva, J., Carvalho, A. S., Ferreira, R., Vitorino, R., Amado, F., Domingues, P., Teixeira, P. & Gibbs, P. A. (2005). Effect of the pH of growth on the survival of *Lactobacillus delbrueckii* subsp. *bulgaricus* to stress conditions during spray-drying. *J Appl Microbiol* 98, 775-782.
- Singh, V. K. & Moskovitz, J. (2003). Multiple methionine sulfoxide reductase genes in *Staphylococcus aureus*: expression of activity and roles in tolerance of oxidative stress. *Microbiology* 149, 2739-2747.
- Sinha, S. C. & Smith, J. L. (2001). The PRT protein family. *Curr Opin Struct Biol* 11, 733-739.
- Sleator, R. D. & Hill, C. (2002). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol Rev* 26, 49-71.
- Solioz, M. & Stoyanov, J. V. (2003). Copper homeostasis in *Enterococcus hirae*. *FEMS Microbiol Rev* 27, 183-195.
- Spano, G., Chiappa, G., Beneduce, L. & Massa, S. (2004). Expression analysis of putative *arcA*, *arcB* and *arcC* genes partially cloned from *Lactobacillus plantarum* isolated from wine. *J Appl Microbiol* 96, 185-193.
- St Amant, D. C., Valentin-Bon, I. E. & Jerse, A. E. (2002). Inhibition of *Neisseria gonorrhoeae* by *Lactobacillus* species that are commonly isolated from the female genital tract. *Infect Immun* 70, 7169-7171.
- Stiles, M. E. & Holzapfel, W. H. (1997). Lactic acid bacteria of foods and their current taxonomy. *Int J Food Microbiol* 36, 1-29.
- Suokko, A., Savijoki, K., Malinen, E., Palva, A. & Varmanen, P. (2005). Characterization of a mobile *clpL* gene from *Lactobacillus rhamnosus*. *Appl Environ Microbiol* 71, 2061-2069.
- Suzuki, K., Benno, Y., Mitsuoka, T., Takebe, S., Kobashi, K. & Hase, J. (1979). Urease-producing species of intestinal anaerobes and their activities. *Appl Environ Microbiol* 37, 379-382.
- Szurmant, H., Nelson, K., Kim, E. J., Perego, M. & Hoch, J. A. (2005). YycH regulates the activity of the essential YycFG two-component system in *Bacillus subtilis*. *J Bacteriol* 187, 5419-5426.
- Takahashi, K., Ishikawa, N., Sadamoto, Y. & other authors (2003). E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics* 19, 1727-1729.
- Tamine, A. Y. & Robinson, R. K. (1999). *Yoghurt Science and Technology*, 2nd edn: Woodhead Publishing.
- Tatusov, R. L., Natale, D. A., Garkavtsev, I. V. & other authors (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 29, 22-28.

- Teusink, B., van Enkevort, F. H., Francke, C., Wiersma, A., Wegkamp, A., Smid, E. J. & Siezen, R. J. (2005). In silico reconstruction of the metabolic pathways of *Lactobacillus plantarum*: comparing predictions of nutrient requirements with those from growth experiments. *Appl Environ Microbiol* 71, 7253-7262.
- Tomita, M., Hashimoto, K., Takahashi, K. & other authors (1999). E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15, 72-84.
- Trchounian, A. & Kobayashi, H. (1999). Kup is the major K⁺ uptake system in *Escherichia coli* upon hyper-osmotic stress at a low pH. *FEBS Lett* 447, 144-148.
- Ueno, Y., Hayakawa, K., Takahashi, S. & Oda, K. (1997). Purification and characterization of glutamate decarboxylase from *Lactobacillus brevis* IFO 12005. *Biosci Biotechnol Biochem* 61, 1168-1171.
- van de Guchte, M., Ehrlich, S. D. & Maguin, E. (2001). Production of growth-inhibiting factors by *Lactobacillus delbrueckii*. *J Appl Microbiol* 91, 147-153.
- van de Guchte, M., Serror, P., Chervaux, C., Smokvina, T., Ehrlich, S. D. & Maguin, E. (2002). Stress responses in lactic acid bacteria. *Antonie Van Leeuwenhoek* 82, 187-216.
- van Rooijen, R. J. & de Vos, W. M. (1990). Molecular cloning, transcriptional analysis, and nucleotide sequence of lacR, a gene encoding the repressor of the lactose phosphotransferase system of *Lactococcus lactis*. *J Biol Chem* 265, 18499-18503.
- Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 60, 407-438.
- Varcamonti, M., Arsenijevic, S., Martirani, L., Fusco, D., Naclerio, G. & De Felice, M. (2006). Expression of the heat shock gene clpL of *Streptococcus thermophilus* is induced by both heat and cold shock. *Microb Cell Fact* 5, 6.
- Vendeville, A., Winzer, K., Heurlier, K., Tang, C. M. & Hardie, K. R. (2005). Making 'sense' of metabolism: autoinducer-2, LuxS and pathogenic bacteria. *Nat Rev Microbiol* 3, 383-396.
- Volker, U., Engelmann, S., Maul, B., Riethdorf, S., Volker, A., Schmid, R., Mach, H. & Hecker, M. (1994). Analysis of the induction of general stress proteins of *Bacillus subtilis*. *Microbiology* 140 (Pt 4), 741-752.
- Volker, U., Maul, B. & Hecker, M. (1999). Expression of the sigmaB-dependent general stress regulon confers multiple stress resistance in *Bacillus subtilis*. *J Bacteriol* 181, 3942-3948.
- Weber, M. H., Klein, W., Muller, L., Niess, U. M. & Marahiel, M. A. (2001). Role of the *Bacillus subtilis* fatty acid desaturase in membrane adaptation during cold shock. *Mol Microbiol* 39, 1321-1329.
- Wei, J., Goldberg, M. B., Burland, V. & other authors (2003). Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* 71, 2775-2786.
- Welch, R. A., Burland, V., Plunkett, G., 3rd & other authors (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci U S A* 99, 17020-17024.
- Weng, M., Nagy, P. L. & Zalkin, H. (1995). Identification of the *Bacillus subtilis* pur operon repressor. *Proc Natl Acad Sci U S A* 92, 7455-7459.

- Whittam, T. S. & Bumbaugh, A. C. (2002). Inferences from whole-genome sequences of bacterial pathogens. *Curr Opin Genet Dev* 12, 719-725.
- Wilkins, J. C., Homer, K. A. & Beighton, D. (2001). Altered protein expression of *Streptococcus oralis* cultured at low pH revealed by two-dimensional gel electrophoresis. *Applied and Environmental Microbiology* 67, 3396-3405.
- Wilkins, J. C., Homer, K. A. & Beighton, D. (2002). Analysis of *Streptococcus mutans* proteins modulated by culture under acidic conditions. *Applied and Environmental Microbiology* 68, 2382-2390.
- Wilkinson, S. P. & Grove, A. (2006). Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins. *Curr Issues Mol Biol* 8, 51-62.
- Willemoes, M. & Kilstrup, M. (2005). Nucleoside triphosphate synthesis catalysed by adenylate kinase is ADP dependent. *Arch Biochem Biophys* 444, 195-199.
- Wilson, D. N. & Nierhaus, K. H. (2004). The how and Y of cold shock. *Nat Struct Mol Biol* 11, 1026-1028.
- Winstedt, L., Frankenberg, L., Hederstedt, L. & von Wachenfeldt, C. (2000). *Enterococcus faecalis* V583 contains a cytochrome bd-type respiratory oxidase. *J Bacteriol* 182, 3863-3866.
- Winter, J., Linke, K., Jatzek, A. & Jakob, U. (2005). Severe oxidative stress causes inactivation of DnaK and activation of the redox-regulated chaperone Hsp33. *Mol Cell* 17, 381-392.
- Winterling, K. W., Chafin, D., Hayes, J. J., Sun, J., Levine, A. S., Yasbin, R. E. & Woodgate, R. (1998). The *Bacillus subtilis* DinR binding site: redefinition of the consensus sequence. *J Bacteriol* 180, 2201-2211.
- Wise, A. A. & Price, C. W. (1995). Four additional genes in the sigB operon of *Bacillus subtilis* that control activity of the general stress factor sigma B in response to environmental signals. *J Bacteriol* 177, 123-133.
- Wonderling, L. D., Wilkinson, B. J. & Bayles, D. O. (2004). The htrA (degP) gene of *Listeria monocytogenes* 10403S is essential for optimal growth under stress conditions. *Appl Environ Microbiol* 70, 1935-1943.
- Wosten, M. M. (1998). Eubacterial sigma-factors. *FEMS Microbiol Rev* 22, 127-150.
- Wouters, J. A., Kamphuis, H. H., Hugenholtz, J., Kuipers, O. P., de Vos, W. M. & Abee, T. (2000). Changes in glycolytic activity of *Lactococcus lactis* induced by low temperature. *Appl Environ Microbiol* 66, 3686-3691.
- Wydau, S., Dervyn, R., Anba, J., Dusko Ehrlich, S. & Maguin, E. (2006). Conservation of key elements of natural competence in *Lactococcus lactis* ssp. *FEMS Microbiol Lett* 257, 32-42.
- Xavier, K. B. & Bassler, B. L. (2003). LuxS quorum sensing: more than just a numbers game. *Curr Opin Microbiol* 6, 191-197.
- Xia, X. (1998). How optimized is the translational machinery in *Escherichia coli*, *Salmonella typhimurium* and *Saccharomyces cerevisiae*? *Genetics* 149, 37-44.

- Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V. & Gordon, J. I. (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299, 2074-2076.
- Yamamoto, Y., Poyart, C., Trieu-Cuot, P., Lamberet, G., Gruss, A. & Gaudu, P. (2005). Respiration metabolism of Group B Streptococcus is activated by environmental haem and quinone and contributes to virulence. *Mol Microbiol* 56, 525-534.
- Yamashita, Y., Takehara, T. & Kuramitsu, H. K. (1993). Molecular characterization of a Streptococcus mutans mutant altered in environmental stress responses. *J Bacteriol* 175, 6220-6228.
- Yoshida, K. I., Fujita, Y. & Ehrlich, S. D. (2000). An operon for a putative ATP-binding cassette transport system involved in acetoin utilization of Bacillus subtilis. *J Bacteriol* 182, 5454-5461.
- Zakharyan, E. & Trchounian, A. (2001). K⁺ influx by Kup in Escherichia coli is accompanied by a decrease in H⁺ efflux. *FEMS Microbiol Lett* 204, 61-64.
- Zamfir, M., Callewaert, R., Cornea, P. C., Savu, L., Vatafu, I. & De Vuyst, L. (1999). Purification and characterization of a bacteriocin produced by Lactobacillus acidophilus IBB 801. *J Appl Microbiol* 87, 923-931.
- Zdobnov, E. M. & Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847-848.
- Zeller, T. & Klug, G. (2006). Thioredoxins in bacteria: functions in oxidative stress response and regulation of thioredoxin genes. *Naturwissenschaften*.
- Zheng, Y., Roberts, R. J. & Kasif, S. (2002). Genomic functional annotation using co-evolution profiles of gene clusters. *Genome Biol* 3, RESEARCH0060.
- Zuber, U. & Schumann, W. (1994). CIRCE, a novel heat shock element involved in regulation of heat shock operon dnaK of Bacillus subtilis. *J Bacteriol* 176, 1359-1363.
- Zuniga, M., Champomier-Verges, M., Zagorec, M. & Perez-Martinez, G. (1998). Structural and functional analysis of the gene cluster encoding the enzymes of the arginine deiminase pathway of Lactobacillus sakei. *J Bacteriol* 180, 4154-4159.
- Zuniga, M., Miralles Md Mdel, C. & Perez-Martinez, G. (2002). The Product of arcR, the sixth gene of the arc operon of Lactobacillus sakei, is essential for expression of the arginine deiminase pathway. *Appl Environ Microbiol* 68, 6051-6058.

Publications annexes

Tolérance et réponse adaptative au stress acide chez *Lactobacillus delbrueckii* ssp. *bulgaricus*

Isabelle GUILLOUARD^a, Eng-Mong LIM^b, Maarten Van de GUCHTE^a,
Christine GRIMALDI^a, Stéphanie PENAUD^a, Emmanuelle MAGUIN^{a*}

^a Unité de Génétique Microbienne, INRA, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France

^b Varoise de concentrés – Coca-Cola Midi, Parc d'activité du Plateau Signe, av. de Berlin,
BP 701, 83030 Toulon Cedex 9, France

Abstract – Tolerance and adaptative acid stress response of *Lactobacillus delbrueckii* ssp. *bulgaricus*. Using several approaches, we studied the acid stress response of different strains of *Lactobacillus bulgaricus*. We showed that after an acid shock, the cell survival varied considerably between strains and that all the studied strains were able to develop an acid adaptive response. For one of these strains, we characterized, using two-dimensional electrophoresis, the modifications in the protein pattern induced by acid adaptation. Out of the 50 proteins up-regulated under the adaptive conditions, 21 were identified after N-terminal sequencing. They belong to different functional categories including stress response and general metabolism. To allow a larger genetic exploration of *L. bulgaricus*, the genome of the type strain ATCC 11842 was sequenced in collaboration with the Genoscope. All together, these genomic and post-genomic data should lead to a better understanding of *L. bulgaricus* physiology and its environmental adaptability.

***Lactobacillus bulgaricus* / acid stress / proteome / genome sequence**

Résumé – En utilisant plusieurs approches, nous avons étudié le comportement de différentes souches de *Lactobacillus bulgaricus* face à un stress acide. Nous avons observé que la survie des cellules à un choc acide variait considérablement entre les souches et que toutes les souches étudiées étaient capables d'initier une réponse adaptative à l'acidité. Pour l'une des souches, nous avons caractérisé, par électrophorèse bidimensionnelle, les changements protéiques intervenant dans la réponse adaptative à l'acidité. Parmi les 50 protéines induites pendant l'adaptation, 21 ont pu être identifiées après séquençage N-terminal. Elles appartiennent à différentes catégories fonctionnelles dont celles des réponses aux stress et du métabolisme général. Afin de permettre une exploration génétique plus large de *L. bulgaricus*, le génome complet de la souche type ATCC 11842 a été séquencé en collaboration avec le Génoscope. Toutes ces données de génomique et de post-génomique nous conduiront à une meilleure compréhension de la physiologie de *L. bulgaricus* et notamment de son adaptabilité environnementale.

***Lactobacillus bulgaricus* / stress acide / protéome / séquence génomique**

* Auteur correspondant : maguin@jouy.inra.fr

1. INTRODUCTION

Les bactéries lactiques jouent un rôle majeur dans l'industrie alimentaire et particulièrement dans l'industrie laitière. En France, la consommation de yaourts et autres laits fermentés est proche de 29 kg par personne et par an (http://www.data.lait.com/fr/ultra_frais/marche.asp). *Lactobacillus delbrueckii* ssp. *bulgaricus* est une bactérie lactique homofermentaire utilisée en co-culture avec *Streptococcus thermophilus* pour la production de yaourts. Outre sa fonction d'acidification du lait par fermentation du lactose en acide lactique, *L. bulgaricus* participe également à l'élaboration des qualités organoleptiques, hygiéniques et probiotiques du produit [12].

Au cours des processus de production du ferment, de fermentation du lait et après l'ingestion des produits fermentés, *L. bulgaricus* est soumis à une variété de stress. L'acidité est un des stress rencontrés par la bactérie aussi bien dans les yaourts (pH 4,3 à 4,5) que dans l'estomac (pH 1,5 à 2) [3]. Le pH optimal de croissance de *L. bulgaricus* est d'environ 6, alors que le pH limite de croissance se situe entre 4,2 et 3,8 suivant les souches et le milieu de culture [10]. L'acidification croissante du milieu, entraînée par la production d'acide lactique, impose donc à la bactérie d'exprimer des mécanismes d'adaptation qui lui permettent de poursuivre sa croissance jusqu'à un pH inhibiteur, puis d'y survivre. Cependant, les mécanismes permettant de résister aux différents stress, et en particulier aux bas pH, sont peu connus chez *L. bulgaricus*.

L'un de nos objectifs est d'étudier les mécanismes responsables de la résistance au stress acide chez *L. bulgaricus*. Nous avons montré une variabilité de la capacité de résistance à un bas pH de différentes souches de *L. bulgaricus*. La caractérisation des protéines induites lors de l'adaptation à l'acidité par électrophorèse bidimensionnelle a été entreprise. Nous avons également débuté un programme de génomique de *L. bulgaricus* qui nous fournira les données génétiques nécessaires à la

compréhension de l'adaptation aux différents stress rencontrés par la bactérie.

2. MATÉRIELS ET MÉTHODES

2.1. Souches bactériennes et conditions de culture

Les souches bactériennes de *Lactobacillus delbrueckii* ssp. *bulgaricus* VI-1004, VI-1007, VI-1010, VI-1019 et VI-1020 (Rhodia Food, Aubervilliers, France et Danone Vitapole, Palaiseau, France) ont été utilisées pour cette étude. Pour chaque expérience, les cultures ont été réalisées en milieu chimiquement défini MPL [2] et incubées à 42 °C. Par dilutions successives, les cultures ont été maintenues en phase de croissance exponentielle précoce ($DO_{600} < 0,3$) pendant au moins 16 h avant les expériences de stress afin d'éviter les modifications de tolérance aux stress liées à la phase stationnaire.

2.2. Test de survie au stress acide

Les cultures en phase exponentielle de croissance sont diluées à $DO_{600} = 0,08$ et cultivées dans 1,5 mL de milieu MPL à pH 6 pendant 1h30. Après centrifugation, les culots cellulaires sont repris dans 1 mL de milieu MPL à pH 3,6. La viabilité est mesurée par étalement sur boîtes de milieu MRS, immédiatement (T_0) et 30 min après le transfert à pH acide (T_{30}). Le taux de survie est égal au nombre d'unités formant des colonies (ufc) à T_{30} rapporté au nombre d'ufc à T_0 .

2.3. Électrophorèse bidimensionnelle

Le marquage et l'extraction des protéines totales, les conditions d'électrophorèse bidimensionnelle, l'analyse des gels et le séquençage de l'extrémité N-terminale des protéines ont été réalisés comme précédemment décrit [9].

2.4. Séquençage

Le génome de la souche de référence *L. bulgaricus* ATCC 11842 a été séquençé en

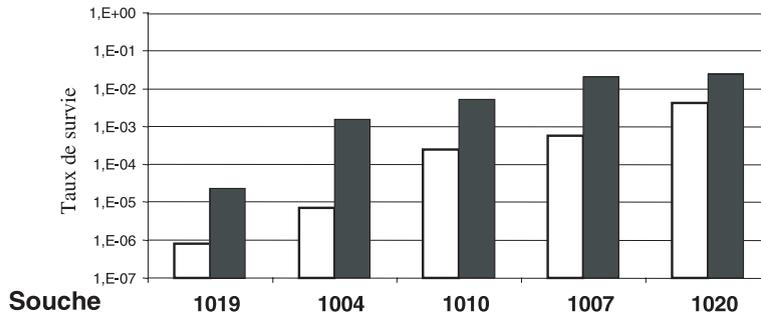


Figure 1. Tolérance et adaptation de différentes souches de *L. bulgaricus* à un choc acide. Les taux de survie (ufc à $T_{30 \text{ min}}$)/(ufc à T_0) au stress acide (30 min à pH 3,6) sont mesurés sur des cellules en phase de croissance exponentielle ($OD_{600} < 0,3$) non adaptées (rectangles blancs) ou adaptées pendant 30 min à pH 4,75 (rectangles noirs).

Figure 1. Acid stress tolerance and adaptation of several strains of *L. bulgaricus*. The survival rate (cfu at $T_{30 \text{ min}}$)/(cfu at T_0) of exponentially growing cells ($OD_{600} < 0.3$) exposed to acid stress (30 min at pH 3.6) was measured for non adapted cells (white bars) or adapted cells (black bars).

collaboration avec le Génoscope (V. Barbe, S. Oztas, S. Manganot, R. Eckenberg et J. Weissenbach) avec une approche par shot-gun.

3. RÉSULTATS ET DISCUSSION

3.1. Tolérance au stress acide de cinq souches de *L. bulgaricus*

Nous avons réalisé une étude comparative de la tolérance au stress acide de cinq souches de *L. bulgaricus*. Des cellules en croissance exponentielle en milieu MPL à 42 °C sont transférées de pH 6 à pH 3,6 pendant 30 min. Leur taux de survie à ce choc acide est alors mesuré (Fig. 1). Dans ces conditions, la souche VI-1020 est 5000 fois plus résistante à un choc acide que la souche VI-1019. Les souches VI-1004, VI-1010 et VI-1007 sont respectivement 10, 50 et 1500 fois plus résistantes à l'acidité que la souche VI-1019. Ces résultats montrent que la tolérance à l'acidité de *L. bulgaricus* est variable selon les souches.

3.2. Comparaison de la réponse adaptative au stress acide de cinq souches de *L. bulgaricus*

Chez plusieurs bactéries, l'exposition à un stress non léthal augmente la tolérance à un stress léthal et permet d'engendrer une réponse adaptative [1, 4, 8, 17]. Cette propriété est également trouvée pour au moins une souche de *L. bulgaricus* qui est capable de mieux survivre à un choc acide léthal si elle a été préalablement exposée à un stress acide modéré [9]. Nous avons comparé la capacité d'adaptation au stress acide des cinq souches de *L. bulgaricus*. Nous avons réalisé le test de survie précédemment décrit sur des cellules en phase de croissance exponentielle adaptées à pH 4,75 pendant 30 min (stress non léthal) puis soumises à un stress léthal à pH 3,6 pendant 30 min. Les bactéries adaptées sont, selon les souches, 5 à 200 fois plus tolérantes au stress acide que les bactéries non adaptées (Fig. 1).

Ces résultats montrent qu'il existe une grande variabilité entre les souches de

L. bulgaricus au niveau de la tolérance à un stress acide. Cependant, toutes ces souches sont capables de mettre en place une réponse adaptative à un stress modéré qui augmente sensiblement leur survie en condition de stress.

3.3. Analyse des protéines induites pendant l'adaptation à l'acidité par électrophorèse bidimensionnelle

Pour obtenir une vision globale des protéines induites pendant l'adaptation à l'acidité chez *L. bulgaricus*, nous avons entrepris une étude protéomique de l'une des cinq souches, VI-1010. Nous avons comparé les profils d'expression des protéines synthétisées à pH 6 (référence) et à pH 4,75 (adaptation).

Le marquage des protéines a été réalisé en milieu chimiquement défini (MPL) [2] contenant de la méthionine [³⁵S] [9]. Trois expériences indépendantes de marquage ont été réalisées et après préparation des extraits protéiques, plusieurs gels 2-D ont été analysés. Afin de repérer les protéines synthétisées lors de l'adaptation, le marquage a été réalisé pendant 15 min après différents temps d'adaptation à pH 4,75. La comparaison des quantités moyennes de protéines synthétisées à pH 6 et jusqu'à 45 min à pH 4,75 montre une synthèse significativement augmentée, d'un facteur supérieur à 2, pour environ 50 protéines (données non montrées).

3.4. Identification des protéines induites pendant l'adaptation à l'acidité par séquençage des acides aminés N-terminaux

Afin d'identifier les protéines induites pendant l'adaptation à pH 4,75, des gels bidimensionnels préparatifs ont été réalisés. Après transfert sur membrane, la séquence peptidique des extrémités N-terminales des protéines a été déterminée par micro-séquençage [16]. Parmi les 50 protéines induites en condition d'adaptation, 21 ont été potentiellement identifiées par séquen-

çage N-terminal. Une fonction leur a été attribuée sur la base d'homologie de séquences avec les protéines déjà connues. Ces protéines peuvent être regroupées en plusieurs catégories fonctionnelles :

- Trois protéines sont fortement induites en condition d'adaptation: GroES, GroEL et DnaK. Ces trois protéines appartiennent à la famille des protéines chaperonnes induites par le choc thermique (HSP) qui sont impliquées dans le repliement et/ou l'adressage aux protéases de protéines dénaturées [9, 20]. L'induction des HSPs en réponse à des conditions de stress variées, notamment le stress acide, a également été décrite chez d'autres bactéries à Gram positif comme *Lactococcus lactis* [6, 7], *Bacillus subtilis* [15] et *Listeria monocytogenes* [5].

- Nous avons observé qu'une des sous-unités de l'ATPase F₀F₁ était surexprimée lors de l'adaptation à l'acidité. Ce complexe protéique expulse les protons hors de la cellule en hydrolysant l'ATP. Son rôle essentiel dans la résistance au stress acide a déjà été établi chez plusieurs bactéries lactiques [13, 14, 19].

- Plusieurs protéines appartenant au métabolisme général sont induites pendant l'adaptation à l'acidité : trois protéines sont impliquées dans le métabolisme carboné dont la 1-phosphofructokinase qui convertit le fructose-1-phosphate en fructose-1,6-bisphosphate. Dix protéines jouent un rôle dans le métabolisme azoté. Parmi elles, la dihydrodipicolinate synthase est induite d'un facteur 10 environ en condition d'adaptation et participe à la biosynthèse de la lysine. Enfin, quatre protéines interviennent dans le métabolisme de l'enveloppe cellulaire. C'est le cas de l'alanine racémase qui participe à la biosynthèse de la paroi cellulaire.

Cependant, plus de la moitié des protéines induites en condition d'adaptation n'ont pas pu être identifiées sur la base des séquences N-terminales en particulier à cause du peu de données génomiques de *L. bulgaricus*.

3.5. Projet de séquençage de *L. bulgaricus*

Le séquençage du génome complet de la souche de référence de *L. bulgaricus* ATCC 11842 a été entrepris en collaboration avec le Génoscope. La séquence génomique a été obtenue en utilisant une stratégie de séquençage par shot-gun suivie de PCRs longue portée (Long range PCR) et de PCR combinatoires (Multiplex PCR [18]). *L. bulgaricus* possède un génome de taille 1,84 Mb et un contenu en GC de 49,8 %. L'analyse du GC skew [(G-C)/(G+C)] [11] donne un profil typique des chromosomes circulaires et permet de positionner une origine de réplication, *oriC*, associée à *dnaA* codant une protéine d'initiation de la réplication, ainsi que le terminus potentiel de réplication.

Chez *B. subtilis* et *Escherichia coli*, l'expression de nombreux gènes impliqués dans la tolérance aux stress est gouvernée par un facteur sigma alternatif (σ^B chez *B. subtilis* et σ^S chez *E. coli*). Aucun orthologue d'un tel facteur sigma n'a pu être identifié chez *L. bulgaricus*. C'est également le cas pour les Streptocoques et Enterocoques dont les génomes ont été séquencés. Ces observations soulèvent d'intéressantes questions quant à la régulation des réponses aux stress chez *L. bulgaricus* et chez les bactéries lactiques.

Les données de séquence nous seront très utiles pour la caractérisation des protéines de stress visibles sur les gels bidimensionnels. Une analyse préliminaire de 40 spots d'intensité variable sur le gel de référence à pH 6, nous a permis d'identifier 36 protéines en couplant la spectrométrie de masse aux données de séquence. Cette forte proportion d'identification nous laisse espérer une détermination rapide de l'ensemble des protéines d'adaptation à pH acide.

4. CONCLUSION

Les données acquises par les différentes approches, qui vont de la génomique à la

post-génomique en passant par le développement et l'utilisation d'outils génétiques, devraient permettre d'acquérir rapidement une meilleure connaissance de la physiologie de *L. bulgaricus*, et notamment de son comportement en réponse aux stress environnementaux. Elles fourniront également une base génétique pour évaluer la biodiversité au sein de l'espèce afin de mieux déterminer le potentiel des souches.

Remerciements: Nous remercions S. Bouddebouze pour son aide technique dans l'élaboration des gels d'électrophorèse bidimensionnelle, C. Chervaux et T. Smokvina pour leur intérêt pour ce travail. Ce travail a reçu le soutien financier de Danone Vitapole (E.-M. Lim, C. Grimaldi, et S. Penaud) et de Rhodia Food (E.-M. Lim).

RÉFÉRENCES

- [1] Belli W.A., Marquis R.E., Adaptation of *Streptococcus mutans* and *Enterococcus hirae* to acid stress in continuous culture, Appl. Environ. Microbiol. 57 (1991) 1134–1138.
- [2] Chervaux C., Ehrlich S.D., Maguin E., Physiological study of *Lactobacillus delbrueckii* subsp. *bulgaricus* strains in a novel chemically defined medium, Appl. Environ. Microbiol. 66 (2000) 5306–5311.
- [3] Conway P.L., Gorbach S.L., Goldin B.R., Survival of lactic acid bacteria in the human stomach and adhesion to intestinal cells, J. Dairy Sci. 70 (1987) 1–12.
- [4] Flahaut S., Hartke A., Giard J.C., Benachour A., Boutibonnes P., Auffray Y., Relationship between stress response toward bile salts, acid and heat treatment in *Enterococcus faecalis*, FEMS Microbiol. Lett. 138 (1996) 49–54.
- [5] Hanawa T., Yamamoto T., Kamiya S., *Listeria monocytogenes* can grow in macrophages without the aid of proteins induced by environmental stresses, Infect. Immun. 63 (1995) 4595–4599.
- [6] Hartke A., Frere J., Boutibonnes P., Auffray Y., Differential induction of the chaperonin GroEL and the Co-chaperonin GroES by heat, acid, and UV-irradiation in *Lactococcus lactis* subsp. *lactis*, Curr. Microbiol. 34 (1997) 23–26.

- [7] Kilstrup M., Jacobsen S., Hammer K., Vogensen F.K., Induction of heat shock proteins DnaK, GroEL, and GroES by salt stress in *Lactococcus lactis*, *Appl. Environ. Microbiol.* 63 (1997) 1826–1837.
- [8] Kim W.S., Perl L., Park J.H., Tandianus J.E., Dunn N.W., Assessment of stress response of the probiotic *Lactobacillus acidophilus*, *Curr. Microbiol.* 43 (2001) 346–350.
- [9] Lim E.M., Ehrlich S.D., Maguin E., Identification of stress-inducible proteins in *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Electrophoresis* 21 (2000) 2557–2561.
- [10] Lim E.M., Lafon A., Dridi L., Boudebouze S., Ehrlich D., Maguin E., Identification de protéines de stress chez *Lactobacillus bulgaricus* par électrophorèse bidimensionnelle, *Lait* 81 (2001) 317–325.
- [11] Lobry J.R., Asymmetric substitution patterns in the two DNA strands of bacteria, *Mol. Biol. Evol.* 13 (1996) 660–665.
- [12] Loones A., Laites fermentés par les bactéries lactiques, in: de Roissart H., Luquet F.M. (Eds.), *Bactéries lactiques*, Loriga, Uriage, France, 1994, pp. 135–154.
- [13] Nannen N.L., Hutkins R.W., Proton-translocation adenosine triphosphatase activity in lactic acid bacteria, *J. Dairy Sci.* 74 (1991) 747–751.
- [14] O’Sullivan E., Condon S., Relationship between acid tolerance, cytoplasmic pH, and ATP and H⁺-ATPase levels in chemostat cultures of *Lactococcus lactis*, *Appl. Environ. Microbiol.* 65 (1999) 2287–2293.
- [15] Petersohn A., Brigulla M., Haas S., Hoheisel J.D., Volker U., Hecker M., Global analysis of the general stress response of *Bacillus subtilis*, *J. Bacteriol.* 183 (2001) 5617–5631.
- [16] Ploug M., Jensen A.L., Barkholt V., Determination of amino acid compositions and NH₂-terminal sequences of peptides electroblotted onto PVDF membranes from tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis: application to peptide mapping of human complement component C3, *Anal. Biochem.* 181 (1989) 33–39.
- [17] Rallu F., Gruss A., Maguin E., *Lactococcus lactis* and stress, *Antonie Van Leeuwenhoek* 70 (1996) 243–251.
- [18] Sorokin A., Lapidus A., Capuano V., Galleron N., Pujic P., Ehrlich S.D., A new approach using multiplex long accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing, *Genome Res.* 6 (1996) 448–453.
- [19] Yokota A., Amachi S., Ishii S., Tomita F., Acid sensitivity of a mutant of *Lactococcus lactis* subsp. *lactis* C2 with reduced membrane bound ATPase activity, *Biosci. Biotechnol. Biochem.* 59 (1995) 2004–2007.
- [20] Yura T., Nagai H., Mori H., Regulation of the heat-shock response in bacteria, *Annu. Rev. Microbiol.* 47 (1993) 321–350.

AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system

K. Bryson, V. Loux, R. Bossy, P. Nicolas, S. Chaillou¹, M. van de Guchte², S. Penaud², E. Maguin², M. Hoebeke, P. Bessières and J-F Gibrat*

Mathématique, Informatique et Génome, INRA, 78352 Jouy-en-Josas Cedex, France, ¹Flore Lactique et Environnement Carné, INRA, 78352 Jouy-en-Josas Cedex, France and ²Génétique Microbienne, INRA, 78352 Jouy-en-Josas Cedex, France

Received February 16, 2006; Revised and Accepted June 20, 2006

ABSTRACT

We have implemented a genome annotation system for prokaryotes called AGMIAL. Our approach embodies a number of key principles. First, expert manual annotators are seen as a critical component of the overall system; user interfaces were cyclically refined to satisfy their needs. Second, the overall process should be orchestrated in terms of a global annotation strategy; this facilitates coordination between a team of annotators and automatic data analysis. Third, the annotation strategy should allow progressive and incremental annotation from a time when only a few draft contigs are available, to when a final finished assembly is produced. The overall architecture employed is modular and extensible, being based on the W3 standard Web services framework. Specialized modules interact with two independent core modules that are used to annotate, respectively, genomic and protein sequences. AGMIAL is currently being used by several INRA laboratories to analyze genomes of bacteria relevant to the food-processing industry, and is distributed under an open source license.

INTRODUCTION

Around 10 years ago, the first prokaryotic genomes were sequenced and annotated (1). Each project represented a milestone for biology and was often carried out by a consortium of laboratories with substantial resources, including adequate bioinformatics support.

The subsequent decade has seen enormous advances in sequencing technologies, with the result that small teams within individual laboratories are now able to sequence

their favorite prokaryotes. Information gleaned from such studies has accelerated the pace of both fundamental and applied biology. However, for many small laboratories, a bottleneck in their progress has been finding bioinformatics expertise to allow them to annotate their genomes of interest. It is now clear that the highest quality annotation arises from manual annotation by experts in the particular organism. So, without bioinformatics support available, one of the key roles of an annotation system is to enable expert biologists to annotate raw genomic data themselves.

Genome annotation is a complex process and involves a number of different dimensions. A substantial aspect is simply coherent data management. Any sequencing project will result in the large numbers of contigs being sequenced and combined, over time, into different assembled versions of the genome. Each assembly requires the application of large numbers of specialist bioinformatics tools, resulting in information about different regions of the genome or proteins expressed therein. The management of such data throughout this complex process is daunting; it is essential that such details are automatically handled by the annotation system.

After basic data management, a second aspect is the overall interpretation of bioinformatics results to form manual annotation. Generally, the expert biologist needs to examine all the automatic analysis and, combined with his/her own knowledge of the organism, form an overall decision on the nature of the different genomic elements, such as genes. This is where bioinformatics expertise is often required, interpreting the results, understanding appropriate thresholds for the different scores, etc. Without bioinformatics support available, it is essential that the analysis system facilitates the annotator to interpret the different analysis results.

Yet another factor to consider is the rapid progress that is also being made in the field of bioinformatics. New tools and databases are constantly under development leading to more accurate predictions. Often the incorporation of complete systems would help annotation, e.g. the inclusion of information

*To whom correspondence should be addressed. Tel: +33 1 34 65 28 97; Fax: +33 1 34 65 29 01; E-mail: Jean-Francois.Gibrat@jouy.inra.fr
Present address:

K. Bryson, Department of Computer Science, University College London, London WC1E 6BT, UK.
M. Hoebeke, Statistique et Génome, Université d'Evry, 91000 Evry Cedex, France.

about genomic sequences overlaid onto KEGG metabolic pathways (which was incorporated into the current system after initial deployment). Annotation systems which are unable to evolve, possibly due to having a fixed architecture or inflexible annotation strategy, will result in annotation which becomes inferior over time.

Considering all these different aspects, it is not surprising that small sequencing teams, without locally available bioinformatics expertise, are left struggling. It is with these objectives in mind, that the AGMIAL consortium [AGMIAL is a French acronym for Analyse de Génomes Microbiens d'Intérêt Agro-alimentaire], consisting of a bioinformatics group collaborating with a number of small sequencing teams, developed the AGMIAL system to satisfy these particular needs.

This article is organized into three parts. We first describe a suitable annotation strategy for prokaryotes. An overview of existing annotation platforms is then provided, based on key features of interest. The AGMIAL system is then described as a practical system, focusing on the aspects above and also drawing together the best ideas from these other annotation systems.

ANNOTATION STRATEGY

Annotation, taken in its broadest sense, is the process of extracting biological knowledge from rather cryptic raw data, the nucleotide sequence. Annotation is a complex task that requires the integration of many data sources, such as the results from bioinformatics analysis tools, data extracted from generic and specific databases, biological knowledge accumulated in the literature over the years and results of genome-wide experiments, such as transcriptomics or proteomics experiments.

Two stages can be identified in the annotation of genomic data (2). The first stage corresponds to a static view of the genome where one describes the fundamental objects that it contains: the genes, their associated *cis*-regulatory regions and the proteins. This stage is necessary but not sufficient to obtain an in depth understanding of the complex relationship between the genome and the biological properties. A more dynamic view must be adopted, i.e. one must consider the multiple ways for genes and proteins to interact so as to create 'functional modules' (metabolic paths, signaling cascades, regulatory loops, etc.), which underlie the biological properties. This second stage constitutes a real challenge that the biology community is beginning to address (3). Transcriptomics and proteomics technologies, together with other high-throughput approaches, are playing a vital role in gaining this systems-level understanding of biology.

The first stage, being the foundation of subsequent analyzes, needs to be addressed properly. With the rapid increase of available genomic data, and the development of new bioinformatics methods to analyze these data a great deal of information relevant to the annotation can be extracted automatically. We think that it is important to go beyond the usual variants of the 'best BLAST hit' strategy and provide the annotators with the most diverse and comprehensive set of data regarding the genes and proteins to be annotated. In the next section we present the different types of information and the corresponding tools we consider pertinent for the

annotation process and we have included in our annotation system.

Identification of genes and other genetic elements

Features of interest in the DNA sequence consist of genes coding for proteins or various types of RNA (tRNA, rRNA, etc.), ribosome binding sites (RBS), terminators, insertion sequences, specific signals (e.g. the CHI site), promoter regions, horizontally transferred regions, repetitions, etc.

A number of bioinformatics tools have been developed to identify these features, in particular much effort has been devoted to gene detection. In prokaryotes, coding sequences represent about 90% of the genome and are not split into exons/introns. These two characteristics allow the development of methods, based for instance on Hidden Markov Models (HMM), that provide very accurate results.

Overall, we think that the detection of genes and other genetic elements for prokaryotes is relatively straightforward, using current tools, with the possible exception of promoters. The subsequent 'Contig Analysis Manager' (CAM) section provides a brief description of the software used within Agmial to identify genetic elements.

Protein functional annotation

The notion of function. Compared to actually identifying a gene, assigning a function to its product is a more challenging task. To begin with, the concept of function is hierarchical, it needs to be described at different levels (4):

- the molecular function that describes the biochemical role of the protein, whether it is a particular enzyme, transporter, repressor, structural protein, etc.
- the cellular function that describes the role of the protein in the cell, e.g. whether it is involved in a particular pathway, a signaling cascade, etc.
- the phenotypic function that describes the effect of the protein on general properties of the organism, e.g. if it is involved in the bacterium gliding ability, in the sporulation process, etc.

In addition, a number of proteins have been shown to possess multiple functions within the cell [the so-called moonlighting proteins (5)]. For instance in *Escherichia coli* a protein which, as a monomer, has a dihydrolipoamide dehydrogenase activity is also found as a subunit of pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase and the glycine cleavage complex (6).

It is also well known that many proteins consist of several domains that have molecular functions of their own. For instance the genetically mobile SH2 domain binds phosphorylated tyrosines. It is found in proteins that needs to recognize other phosphorylated proteins, for instance within a signaling cascade.

Modular aspect and intrinsic properties of protein sequences.

The analysis of a protein sequence should always start with the determination of its underlying substructure, i.e. the identification of the different modules it is made of. This is important for two reasons. The first one is that some regions, such as low complexity regions, can alter subsequent homology searches by causing spurious resemblances between unrelated proteins. The second reason is that ignoring the modular

aspect of proteins is the cause of a well known annotation error whereby the function of a protein is transferred to another one that only shares one module, not related to the general function of the protein (7).

One important type of protein module is the globular domain, although a number of other types also exist. These include transmembrane segments and signal peptides, together with regions of structural disorder, low complexity and coiled-coil. A number of databases and tools exist to predict the presence of globular domains, such as genetically mobile domains (8).

Besides modular aspects of proteins, other global properties, such as the molecular mass and isoelectric point are important, particularly in relation to proteomics studies. Protein properties that can be deduced from codon usage are also informative, for instance those concerning protein abundance and whether the gene is in an atypical region of the genome, possibly indicating horizontal transfer.

Homology search. Homology search techniques are the cornerstone of functional annotation. It is well known that methods based on the comparison of a single sequence, such as BLAST (9) or FASTA (10) become inefficient when they reach the 'twilight zone', about 25–30% sequence identity. Remote homologue detection can be improved if one uses multiple sequence alignments, either building them on the fly from the query sequence like PSI-BLAST (11) or employing protein family alignments (12) to create a statistical model representative of the family with a HMM.

For remote homologues whose sequences have strongly diverged, and can no longer be detected by sequence comparison techniques, it is possible to search for motifs or functional signatures (13). This is a powerful technique but it requires the motif residues to be more or less contiguous in the sequence. If this is not the case, one can use fold recognition techniques that are based on 3D structure conservation property, to detect remote homologues (14).

Genomic context information. The techniques discussed above consider proteins as isolated entities. With the availability of an increasing number of complete genomes, it now seems appropriate to consider the context of a gene between different genomes to help elucidate its function (15).

Techniques based on the genomic context use the colocalization of genes at various levels of physical proximity. They can be used to obtain information about protein function from chromosomal context but, in addition, they can also provide clues about the functional interactions between proteins thereby providing a first step towards cellular process annotation. Three major types of technique exist: gene fusion, gene neighboring and phylogenetic profiles.

Genomic context techniques provide links, different from the link provided by the homology relationship, between proteins of the genome. When this information is combined with data coming from homology search techniques it permits one to gain insight about protein function. It must be noted that this information is far from being marginal. It has been shown for *E.coli* K-12 that genomic context techniques allow one to obtain information for a fraction of genes in the genome similar to the fraction for which homology relationship can be found by sequence comparison methods (16).

Subcellular localization. The subcellular localization is an important practical piece of information, in particular in view of subsequent experiments with the organisms. Four localizations can be defined for Gram positive bacteria: cytoplasm, cytoplasmic membrane, cell wall and exterior (for secreted proteins) and five localizations for Gram negative bacteria: cytoplasm, cytoplasmic membrane, periplasm, outer membrane and exterior.

Different techniques are available to predict protein localization. The most straightforward is based on homology. If a protein is homologous to a protein whose localization is known one simply assumes it has the same localization. The second technique involves the identification of the biological mechanism responsible for the addressing of the protein to its localization, for instance signal peptides for secreted proteins, segments characterized by a high content in apolar residues for membrane proteins. The last technique is based on the amino acid composition of protein sequences. This composition shows a slight but detectable bias according to the localization. These techniques can be combined and weighted accordingly to improve the overall localization prediction (17).

Cellular process annotation

Relatively few bioinformatics tools are available to help biologists in this second stage of the annotation. As we mentioned above, genomic context techniques can be considered as a first step towards the study of protein interactions in the cell. High level functional modules must be studied with a number of large scale experiments. It is therefore important to facilitate the integration of annotation data and functional genomics databases.

EXISTING ANNOTATION PLATFORMS

The desirable features of an annotation platform as listed in the introduction lead to the following characteristics for the developed tool:

- technical points must not concern human experts, in particular the implementation of the annotation strategy must be fully automated;
- human interaction with the results provided by the system must be made as easy as possible.

The latter point can be best implemented through the use of interactive graphic interfaces. Graphic interfaces must allow the visualization of different features at the DNA or protein level, make the bioinformatics analysis results easy to consult. They must provide powerful and flexible ways to manipulate the underlying mechanisms used to query and combine the data, results and annotations, and keep a log of the modifications carried out on the annotations. In addition several annotators must be able to work in parallel on the same data.

We consider that the term 'annotation' should be understood in its broadest meaning, as described in the previous section. This requires the implementation and maintenance of a comprehensive annotation strategy, permitting one to extract as much relevant information as possible from the

available data. As a consequence, it must be easy to integrate new or improved bioinformatics tools and databases into the system when required, creating a 'federation' of tools cooperating together for the purpose of annotating new genomes. The resulting system must be highly modular and robust, based on well tested computer science technologies.

To cope well with most sequencing projects the system must be able to work with draft sequences, in particular it must carry forward, automatically, manual annotations from the previous batch to the new one.

Finally, we firmly believe in the value of open source developments for promoting bioinformatics research in the community. We intend to distribute our system under a GNU Public License and so we only wish to integrate open source, or freely distributed, software into the system.

When we started the AGMIAL project we carried out an analysis of the most salient features of the tools then available. Since then a number of new systems have been developed. Table 1 presents an overview of some of these systems. The list is not meant to be exhaustive, yet, representative of the different features of the tools developed (note that in this table we do not consider systems developed by commercial companies).

This is not the place to give a detailed analysis of the characteristics of all these systems. In particular, they clearly differ both in the technical and conceptual solutions that have been adopted by the developers. Conceptual solutions concern, for instance, the representation of the biological data and the architecture underlying the system: pipeline, workflow, multi-agent system, various types of interacting 'layers', etc. Although conceptual solutions do have an influence on the system capabilities, in the following, we just

restrict ourselves to the description of those features that are directly pertinent to the annotators, along the lines described at the beginning of this section.

Features of interest for our purpose are the following. Manual annotation indicates whether the platform is designed to allow human experts to validate the results and provide the final annotation. Automatic processing of data refers to the capability of the platform to carry out bioinformatics analysis without human intervention. Graphics interface indicates whether the interaction between the annotators and the system takes place mostly through the use of a graphical interface, in other words, whether the interface is really central to the process of manual annotation. Thus, the graphical interface must allow the user to inspect the nucleic sequence and associated features, as well as various results coming from bioinformatic tools used during the analysis. This interface must be interactive, allowing the annotator to modify the data presented, to add new information and to provide mechanisms to perform various searches and comparisons. Collaborative annotation refers to the possibility for several teams of annotators to work on the same genome and possibly, when the genome is published, for biologists browsing the data to add new information or propose corrections. Reasoning capabilities refers to the ability of the system to analyze automatically the results produced by the different bioinformatics tools and predict a particular function in consequence. This criterion might appear in contradiction with the emphasis we put on the central role of the human expert during the annotation process. In fact, as we will discuss later, these 'reasoning' capabilities are not intended to replace human experts but to assist them in their task. Assessment of annotation describes whether the platform provides some means of

Table 1. Characteristics of some annotation platforms

Method	Reference	Organisms	Graphic interface	Automatic processing of data	Manual annotation	Collaborative annotation	'Reasoning' capabilities	Annotation assessment	Availability
MAGPIE	(40)	Prokaryote	No	Yes	Limited	No	Yes	No	Code available
GENOTATOR	(41)	Eukaryote	Yes	Yes	possible	No	No	No	Code available
GAIA	(42)	Eukaryote	Yes	Yes	Limited	No	No	No	Web use
IMAGENE	(43)	Prokaryote	Yes	Possible	Yes	No	No	No	Available ^b
GENEQUIZ	(44)	Both ^a	No	Yes	No	No	Yes	Yes	Web use
ARTEMIS	(24)	Both	Yes	No	Yes	No	No	No	GPL ^c
M-AGENTS	(45)	Virus	No	Yes	No	No	No	No	Web use
PEDANT	(46)	Both	Yes	Yes	Possible	No	No	No	Web use
ENSEMBL	(47)	Both	No ^d	Yes	No ^e	Yes	No	No	GPL
APOLLO	(48)	Both	Yes	No	Limited	Yes	No	No	GPL
OTTER	(49)	Both	Yes	No	Yes	Yes	No	No	Code available
RICEGAAS	(50)	Eukaryote	Yes	Yes	Limited	No	No	No	Web use
GENQUIRE	(51)	Both	Yes	No	Yes	No	No	No	GPL
ATUGC	(52)	Prokaryote	No	Yes	No	No	Yes	No	No
GENDB	(53)	Prokaryote	Yes	Yes	Yes	No	Yes	No	GPL
ASAP	(54)	Both	Yes	No	Yes	Yes	No	No	Web use
SABIA	(55)	Prokaryote	No	Yes	No	No	No	No	Code available
MANATEE	(see note ^f)	Both, virus	Yes	Yes	Yes	No	No	No	GPL
MAGE	(56)	Prokaryote	Yes	Yes	Yes	Yes	No	No	Web use
AGMIAL		Prokaryotes	Yes	Yes	Yes	Yes	No	No	GPL

See the text for a detailed definition of the column headings.

^aGeneQuiz carries out protein sequence analysis only. It has been mostly used to re-annotate prokaryotic genome.

^bRequires ILOG licensed libraries.

^cGNU public license.

^dProvided by APOLLO.

^ePerformed with OTTER.

^f<http://manatee.sourceforge.net>.

quantifying the confidence the annotators have in their annotation. Availability specifies the distribution status of the tool.

Not all the tools described in Table 1 are annotation platforms according to conventional meaning of term. GENQUIRE is an annotation browser/editor, GENEQUIZ is a tool exclusively devoted to protein analysis, APOLLO and OTTER are subsystems of ENSEMBL permitting respectively to browse/visualize DNA features and to manually annotate. The term 'graphic interface' covers different realities: from static Web pages (MAGPIE and GENEQUIZ), to interactive applets (GAIA) upto complete, independent, applications (IMAGENE, ARTEMIS, GENQUIRE and APOLLO). These different systems first introduced some of the key features which are described above. MAGPIE was built around two PROLOG daemons allowing some reasoning on the data to be performed. GENEQUIZ generalized this feature with its module GQreason, it was also the first system to provide an assessment of the annotation accuracy. GENOTATOR was the first platform to provide a graphical browser and to stress the importance of visualization. IMAGENE introduced the notion of 'strategy', i.e. a series of elementary tasks strung together to accomplish a particular goal, (such as predicting CDS) and one of the strengths of this system was to allow the user to define, build and manipulate these strategies. More recently, genomic institutes have integrated a number of these features into comprehensive annotation platforms, e.g. PEDANT (Institute for Bioinformatics, MIPS), ENSEMBL (Sanger Centre), MANATEE (TIGR) and GENDB (Center for Genome Research).

To summarize Table 1, differences between platforms lie in the way genomic data are processed, the spectrum going from completely automatic systems, such as GENEQUIZ to annotation browser/editors, such as GENQUIRE, the relative weighting between protein-oriented versus DNA-oriented analysis, the annotation strategy implemented through the set of bioinformatic tools used, the emphasis put on the role of human experts and collaborative annotation, and the use or not of some formal representation of biological knowledge (ontology/hierarchical classification of functions).

Several classifications are currently used for (microbial) genome annotation, for instance Riley's functional hierarchy (18), Gene Ontology (19) and MIPS Functional Catalogue (20). It is very important, if one wants to fully benefit from the accumulated data, that a broad agreement emerges from the community regarding the description of biological concepts. From a practical viewpoint, the use of a functional classification enforces the use of a controlled vocabulary that facilitates the comparisons between different annotated genomes.

AGMIAL PLATFORM

Overview

Key to the system are two managers called the protein analysis manager (PAM) and the CAM. The PAM has overall responsibility for managing and analyzing proteins. The CAM is more specialized and is responsible for managing, analyzing and consistently updating batches of assembled contigs within a particular genome sequencing project. Both

are independent components, which are complete applications in themselves. They engage in a long term commitment in notifying each other of changes in their respective views of data thus cooperating in the overall task of genome annotation.

Both managers share the same internal architecture. They provide:

- a web interface so users can administrate projects and manually annotate;
- a suite of bioinformatics methods to analyze the data;
- an underlying relational database for storing proteins or contigs respectively, results of analysis tools and annotations;
- a mechanism allowing managers to communicate and exchange relevant data when required.

More specific detail of both managers is given in the following sections.

The CAM

Data input and output. Currently the platform is geared toward annotating new genomes and contig sequence data are imported as nucleotide Fasta format files. For re-annotation, where one needs to compare the old annotation with the updated one, we have included a data import mechanism from EMBL/GenBank format files. To publish the data, which requires it to be deposited in public collections, such as GenBank or EMBL, the system outputs in GenBank/EMBL format files, or tabulated flat files used as input by SEQUIN (for submission at genomic centers).

Bioinformatics tools. Bioinformatics tools that have been incorporated in the CAM are listed in Table 2.

For gene detection and classification we use an in-house system called SHOW (Sequence HOMogeneity Watcher) that is based on HMMs. Gene classification relies on a partitioning of the DNA sequence into regions having a homogeneous composition in words of variable length (21). In the HMM, the detection of genes takes into account: (i) the start and stop codons, (ii) the phased composition of the coding sequence, (iii) the presence of a RBS upstream of a gene and (iv) the possibility for genes to overlap. Parameters of the model are estimated automatically from the DNA sequence to be analyzed, the method does not need special training sets, it adapts itself to the corresponding genomic data. One feature of this approach is that the use of fixed parameters is

Table 2. Bioinformatics tools integrated into the CAM

Method	Description	Reference
SHOW	Gene detection using a hidden Markov models	http://www-mig.jouy.inra.fr/ssb/SHOW/
tRNAScan	Detection of tRNA sequences	(22)
rRNAScan	Detection of rRNA sequences.	
PETRIN	Detection of terminator sequences	(23)

minimized, such as minimum allowed gene length. The system is thus less biased by arbitrary cut-offs and, for instance, is able to detect very short genes. Some of these putative short genes are interesting and we have on-going collaborations to verify them experimentally. Tests of the method on 839 genes of *E.coli* whose products are experimentally known (EcoGene dataset) have shown that SHOW is able to correctly identify the precise limits of 93% of these genes.

Other genetic elements which are determined include tRNAs, rRNAs and terminator sequences. Genes coding for tRNAs are predicted with tRNAscan (22). Ribosomal nucleic acid sequences are detected using an in-house system called rRNAscan, which essentially does a BLASTN (9) search against a database containing known rRNA sequences. Finally, terminator sequences are detected using PETRIN (23).

The results of each analysis method are represented in terms of nucleic acid features, similar to GenBank features but with extended qualifier lists to cater for data management and provenance. Annotators are free to define their own features and qualifiers to better describe the characteristics of the nucleic sequence, if they feel the need. However, these user-defined features and qualifiers will not be included in the canonical GenBank/EMBL file produced for genome publication.

The translations for coding regions, or CDS features, are automatically generated using the appropriate genetic code. CAM then sends these translated proteins to PAM, whereupon they are automatically analyzed and an overall functional description is established (see 'The PAM' section below).

An important aspect which the CAM needs to handle is when a newly assembled batch of contigs is added to the system. Manually edited annotations from the previous batch are automatically forwarded to the new batch. This allows the user to start a project with an unfinished sequence and to have the different annotations transmitted to new batches.

User interface. Annotators interact with the system exclusively through graphic interfaces. These interfaces provide mechanisms to retrieve and visualize the data stored in the databases, to carry out complex searches on the data and, if required, to edit these data.

Currently CAM has three different graphic interfaces that provide views of genomic data at different scales. The first one is an Artemis client allowing annotators to visualize features on the DNA sequence. Artemis is a fully fledged genome analysis and annotation editor developed at the Sanger Centre (24). We have interfaced it with CAM, permitting the annotator to use its functionalities. For instance it can be employed to browse and to edit parts of contigs within a particular project.

In addition to Artemis, we have recently developed a new graphical interface based on an in-house software package called MuGeN (25) (see Figure 1). Our interest in MuGeN stems from the fact that it is a tool built for navigating through multiple annotated genomes. It thus facilitates cross-genome comparison. In particular for genome re-annotation, it allows the new annotation to be compared side-by-side against the original annotation.

The last interface consists of the visualization software CGView (26). This software generates high quality,

zoom-able maps of circular genomes. It provides very useful views able to summarize various features and properties of the complete genome. These views are particularly suitable as illustrations in publications. We have interfaced CGView with CAM so that it can extract data from the relational database and present it in a circular context (see Figure 1).

The PAM

Data input and output. The PAM can be used, independently of the CAM, to analyze a batch of protein sequences. In this configuration, it imports protein sequences as Fasta format files. As part of the annotation platform, translated protein sequences are directly provided by the CAM using the exchange mechanism between the managers. Annotated proteins can be exported as Uniprot format files.

Bioinformatics tools. We have integrated into PAM a number of tools required to implement the annotation strategy described previously (Table 3).

The first set of tools concerns the partitioning of protein sequences into domains. Non globular domains exhibiting a particular composition in amino acids, such as low complexity region (27), transmembrane segments (28), disordered regions or having some internal organization, such as coiled-coils regions (29), signal peptides (30) are searched for. Globular domains are detected using tools integrated in InterProScan (31).

The second set of tools concerns homology searches. Methods to detect protein family domains in InterProScan, such as PFAM (12) or TIGRFAMS (32), belong to this category. We use PSI-BLAST (11) with a number of databases. Some are generic, such as SWISSPROT, and others are specific, for instance, collections of protein sequences belonging to related organisms chosen by the biologists in charge of the project. Protein sequences of the genome are clustered into paralog families. We integrated the CDD collections (33), that makes use of RPS-BLAST, principally to have access to the cluster of orthologous genes (COG) (34) dataset. Multiple sequence alignments resulting from these analyses can be inspected and edited with Jalview (35) a Java multiple sequence editor. We perform a search for protein family signatures or motifs with PROSITE (13) and PRINTS (36) that are both available in InterProScan. Finally, the fold recognition method FROST (14), can be used to carry out a search for remote homologues.

Protein subcellular localization is predicted using PSORTb (17). PSORTb is based on a multiple classification approach employing different computation techniques to analyze various features: signal peptide, transmembrane helices, homology to proteins of known localization, amino acid composition and motifs.

We are currently in the process of integrating genomic context information. We have carried out a cross comparison of all available microbial genomes (268 at the time of writing) and compiled the required information to analyze gene fusion and gene neighborhood conservation. All the results are stored in a relational database with other relevant information. To allow biologists to fully, and easily, exploit these data we are presently designing an efficient user interface.

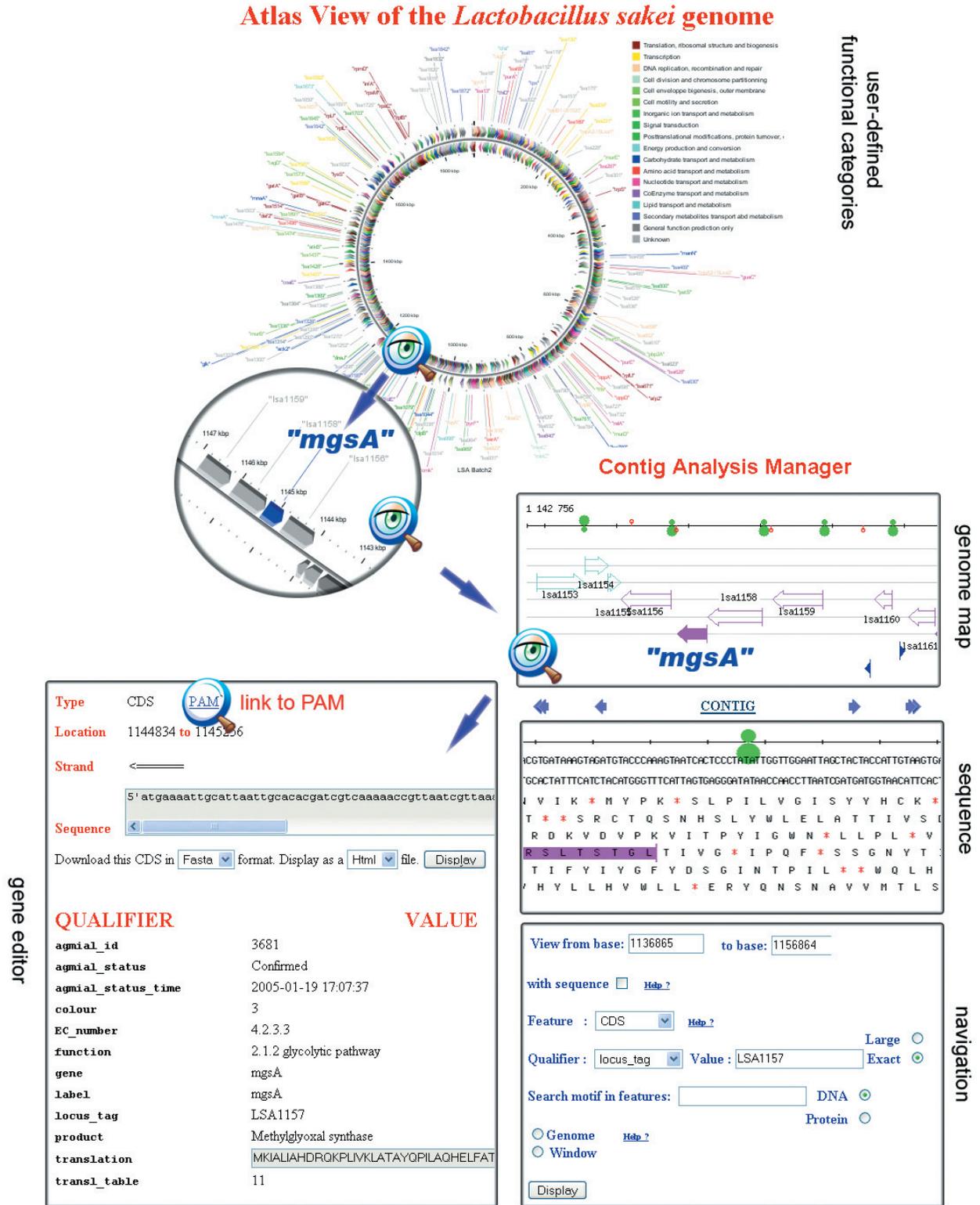


Figure 1. Views of the DNA sequence at different scales. The upper part of the figure represents an atlas view of the genome obtained with CGView. One can zoom on a particular region of this map, for instance on the area containing the *mgsA* gene: a methylglyoxal synthase that belongs to the glycolytic pathway. Clicking on this gene will open the MuGeN interface showing its genomic context (genome map frame). It is possible to zoom on this representation to see the DNA sequence and the translation in the six reading frames (sequence frame). The green symbol represents the RBS, the gene sequence is colored as in the previous view. The navigation window allows one to move along the genome, either by entering a range of base numbers, or by looking for a feature with a particular qualifier or by specifying a DNA or protein motif to be searched for in the current window or in the complete genome. The window at the lower left of the figure shows the gene editor. Most fields are automatically filled, in particular the gene annotation qualifiers since, in general, CDS annotation is performed in PAM and then updated in CAM (see Figure 4). Clicking on the link to PAM, indicated by the magnifying glass, will lead the annotator to the PAM interface shown in Figure 2. For clarity the Artemis interface is not shown on the figure.

Table 3. Bioinformatics tools integrated into the PAM

Method	Description	Reference
Methods to determine sequence intrinsic properties		
pI	Isoelectric point and molecular mass	
SEG	Detection of low-complexity regions	(27)
COIL	Detection of coiled-coil structures	(29)
SIGSEQ	Detection of signal peptides	(30)
MEMSAT	Transmembrane segment prediction	(28)
Homology search methods		
RPS-BLAST	Reverse position specific BLAST	
PSI-BLAST	Sequenced-based homology search	(11)
FROST	Fold recognition method for detecting remote homologues	(14)
Miscellaneous methods		
PSORTb	Prediction of subcellular localization	(17)
Jalview	Multiple sequence alignment editor	(35)
InterProScan	Integrated protein motif detection, including the following software:	(31)
Motif and functional signature detection methods		
ProfileScan	Find motifs using profiles	(57)
ScanRegExp	Find motifs using regular expressions	(58)
FPrintScan	Find multiple motifs	(59)
Domain detection using HMM methods		
HMMSmart	Genetically mobile domains from SMART	(8)
HMMPFam	Protein family domains from PFAM	(12)
HMMTigr	Protein families from TIGR institute	(32)
HMMPIR	Protein families from PIR	(60)
HMMPanther	Protein families subdivided into functionally related subfamilies	(61)
SuperFamily	Proteins of known 3D structure	(62)
Gene3D	Protein families in complete genomes	(63)
Method to split protein sequences into domains		
BlastProDom	Defines domains in protein sequences	(64)
Methods to find sequence intrinsic properties		
SignalPHMM	Prediction of signal peptide	(65)
TMHMM	Prediction of transmembrane helices in proteins	(66)

User interface. Results provided by the above tools are displayed as web pages that annotators can consult with their favorite web browser (see Figure 2). Results are organized in sections: general properties, homology results, feature results, paralog results, etc. The ordering and appearance of these sections can be parameterized by the user. Multiple cross references to the databases used by the different tools (see Table 4) and to the tools themselves exist allowing annotators to browse relevant information.

Homology search methods provide often a large number of homologous proteins. To unravel the complex evolutionary relationship between these proteins and the query protein it is useful to inspect the multiple sequence alignment and to consider the resulting phylogenetic tree. To carry out this task we have interfaced the Jalview multiple sequence editor that allows annotators to browse, and manipulate in a number of ways, the multiple alignment. In addition Jalview can draw the corresponding phylogenetic tree and also cluster the sequences using a principal component analysis technique. This provides a very effective tool for the annotators in their task of assigning a function to the query protein.

At the cellular level it is important to describe pathways responsible for cellular processes. To help annotators studying these pathways and the proteins involved therein, we have developed a graphical tool built on a relational model of the KEGG database (37). It allows annotators to automatically

visualize genome proteins involved in specific metabolic pathways. When a related genome is available, this tool also permits an easy comparison of the proteins involved in the same pathways (see Figure 3). Using this tool it is straightforward, for a specific pathway, to identify proteins that are only present in one of the two genomes and proteins that are common to both. This feature allows annotators to quickly pinpoint, in the studied pathway, the major differences and similarities between the two organisms.

System usage

The initial stage of the annotation process is entirely automatic. The annotation procedure starts when an authorized user loads in CAM the first batch of assembled contigs for the sequencing project, possibly consisting of hundreds of short contigs. As the genome sequencing and assembly continues, new batches of fewer, longer, contigs will be added to the project.

The contigs are automatically analyzed by the tools in CAM and the results, together with the initial data, are stored in the CAM relational database. CDS features are then translated using the appropriate genetic code and the corresponding protein sequences are sent to PAM whereupon they are automatically processed by the various tools and an overall function description is proposed when possible (for the time being, based on a very simple, automatic, analysis of the homologue list). The corresponding results and data are stored in the PAM relational database.

Human experts take part in the annotation process after the completion of the first stage. Using the interfaces described above they are able to consult the results, visualize the features on the DNA sequence, carry out various types of searches and edit the data stored in the database. In AGMIAL both the contig and protein views are integrated. It is thus possible, while examining genes within a particular DNA region, to switch to the PAM interface to consult the available information for the corresponding proteins. Naturally, the converse is also true, the annotator that examines results for a given protein can visualize the DNA region around the corresponding gene with a simple click of the mouse. In a similar fashion, the annotator, while browsing the general view of the genome provided by CGView, can also click on a particular gene to move to a detailed view of the DNA region around it, in MuGeN.

Genome proteins undergo a change of status during the annotation process. When CAM first transfer a protein sequence to PAM the status is set to 'original'. Following the automatic analysis of the sequence, PAM may suggest a function for the protein whose status becomes 'automatic'. After pondering the different results and data available, annotators may confirm the assigned function by changing the status of the prediction from 'automatic' to 'confirmed'. Alternatively, they may decide to modify and update the functional description before changing it to 'confirmed'.

It is important to note that both managers store every description and status change a protein goes through. So the detailed history of a protein's annotation is kept and displayed by the web interface. In this way PAM provides secure, persistent storage of collections of protein sequences within projects.

Figure 2. The right part of the figure shows the results of the different bioinformatic methods applied to the sequence of MgsA. Not all result sections are shown here. In the 'homology' section, checking the boxes on the left of the homologous sequences and then clicking on the link to Jalview, below, will show the multiple alignment of the selected sequences and the corresponding phylogenetic tree. The left part of the figure shows the annotation window where the protein annotation is performed. Information entered in this section is forwarded to CAM, the system always makes sure that both managers are synchronized. The bottom of this window shows the annotation history. The link to CAM at the top will lead the annotator back to the CAM interface (MuGeN interface, see Figure 1). The link to PAREO (our relational version of the KEGG database) near the 'EC number' box will lead the user to the KEGG interface shown in Figure 3.

Table 4. Databases used by the PAM tools

Name	Description	Reference
UNIPROT	Protein sequences and functions	(67)
KEGG	Molecular interaction networks	(37)
CDD	Conserved domain database	(33)
PDB	3D structures database	(68)
SCOP	Protein 3D domain database	(69)
PROSITE	Functional motifs and profiles derived from SWISSPROT	(70)
PRINTS	Manually derived functional motifs	(36)
SMART	Motifs of genetically mobile domains	(8)
PFAM	Protein family domains	(71)
TIGRFAMs	Similar to PFAM	(72)
SUPERFAMILY	Families of proteins of known 3D structures	(62)
GENE3D	Protein families and domain architectures in complete genomes	(63)
PANTHER	Protein families subdivided into functionally related subfamilies	(61)
PRODOM	Automatically generated protein domain families	(64)

For annotators, it is critical that protein annotations in PAM and the corresponding CDSs in CAM are synchronized. In order to achieve this, both managers establish a dialog. Actions of annotators on the data on any manager automatically result

in an exchange of information between this manager and its counterpart so the annotation is kept consistent and up-to-date on both sides. For instance, each time an annotator edits protein features using the PAM web interface, PAM sends the modifications to CAM which updates the corresponding CDS in its database. Similarly, whenever a user edits a CDS feature with CAM, e.g. modifying the start of a gene, CAM sends a message to PAM which updates its database accordingly for the corresponding protein. When the modification results in a new protein the whole battery of PAM tools is automatically applied to the new sequence. The old gene and protein are kept in the databases but marked as disabled. The interaction between the two managers and the user is shown in Figure 4.

APPLICATIONS

Annotation tool

The AGMIAL system is currently being employed by several laboratories at INRA to annotate different genomes of interest, for instance, *Lactobacillus sakei* (38), *Lactobacillus bulgaricus* (39), *Flavobacterium psychrophilum* (submitted), *Staphylococcus xylosus*, *Propionibacterium freudenreichii*,

PAREO

Glycolysis / Gluconeogenesis comparison between *Lactobacillus plantarum* and *Lactobacillus sakei*

[Home](#) | [Metabolic classification](#) | [Enzyme classification](#) | [Compound classification](#)

Display this pathway for Reference pathway

submit

Enzymes in *Lactobacillus plantarum* *Lactobacillus sakei* both.

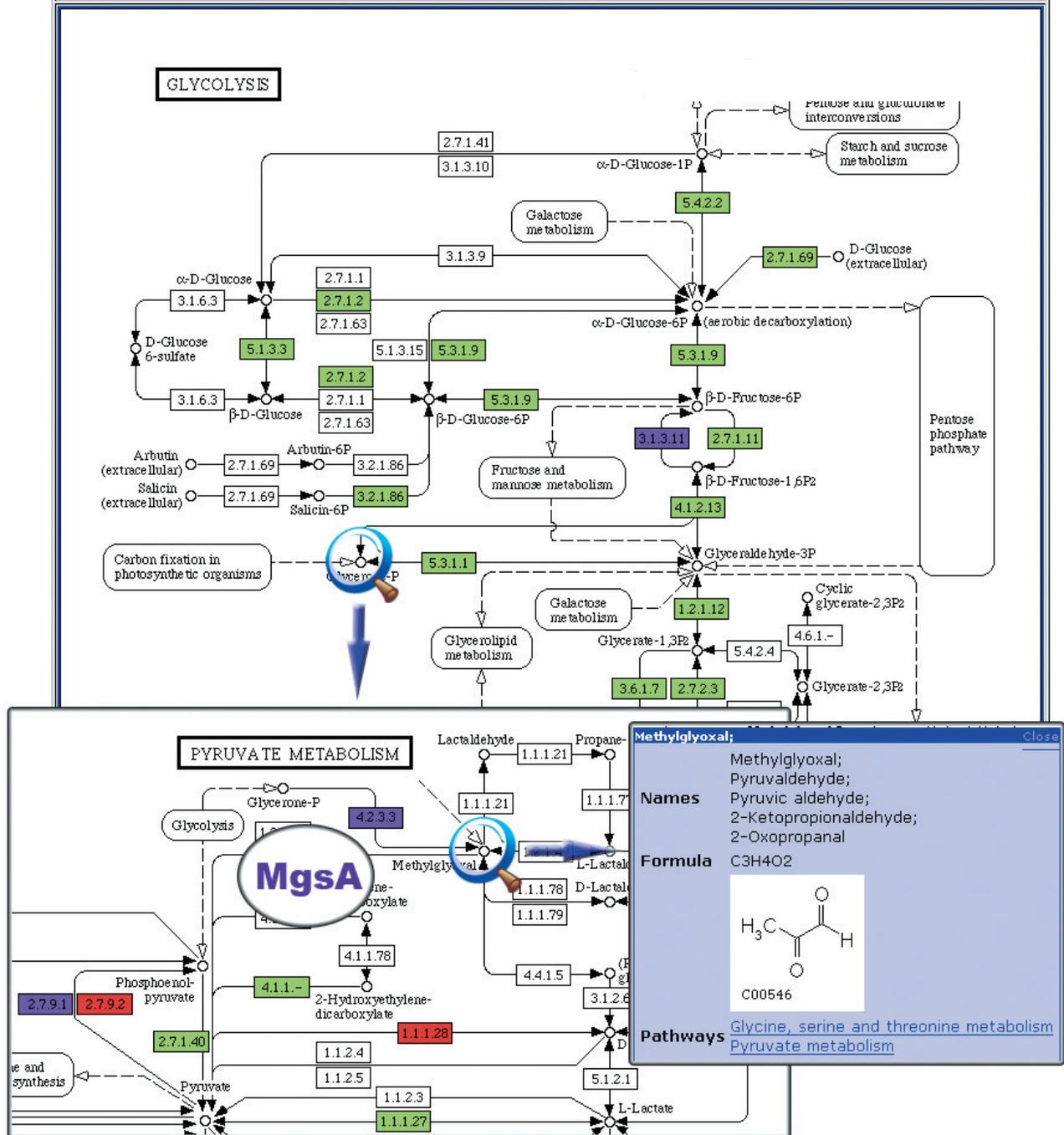


Figure 3. This figure shows both glycolysis and pyruvate metabolism pathways for *Lactobacillus plantarum* and *L. sakei*. As indicated in the legend at the top of the figure, enzymes that are only found in *L. plantarum* and *L. sakei* are colored respectively in red and purple. Enzymes found in both organisms are colored in green. The magnifying glasses are used to indicate the role of MgsA in these pathways. This enzyme appears to be involved in a methylglyoxal bypass (reversible reaction) of glycolysis in *L. sakei*. The figure illustrates well the major difference in glycolysis in *L. plantarum* and *L. sakei*. The bottom right box shows the product of the reaction catalyzed by MgsA. A detailed account of *L. sakei* energy production pathways contributing to meat adaptation can be found in ref. (38).

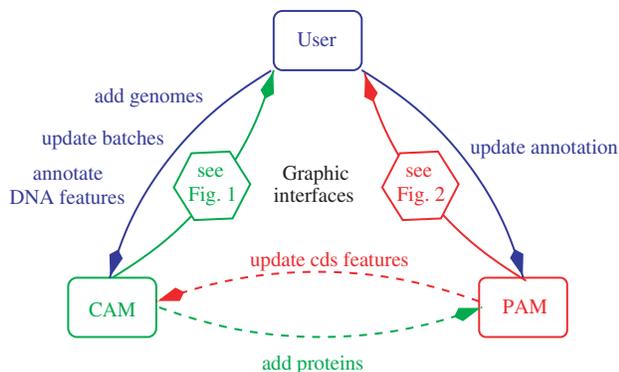


Figure 4. Dashed arrows represent automatic processes between managers, solid arrows represent human interaction with the managers. Graphic interfaces are described in Figures 1 and 2.

Arthrobacter arilaitensis and the strains JIM8777 and JIM8780 of *Streptococcus salivarius*.

Re-annotation tool

Besides using the AGMIAL platform to annotate newly sequenced genomes, a number of INRA groups were interested in re-annotating already published genomes, in particular when different strains of some organism of interest, or closely related organisms, were known (e.g. *Enterococcus faecalis*, *Enterococcus faecium*, *Lactococcus lactis*, *Bacteroides thetaiotaomicron*, *Streptococcus thermophilus*). The fact that the genomes are processed by the same set of tools and stored under the same relational schema facilitates considerably the comparison of the different strains, or related organisms. It also provides a good framework for data mining techniques or other bioinformatics methods.

Genome database

Once the genomes are annotated or re-annotated the platform can be used as a model organism database. We identify two potential groups of users:

- Biologists who will be able to visualize genome features, browse and retrieve corresponding annotations using the graphical interfaces provided.
- Bioinformatics groups that can take advantage of the availability of the source code, the use of computer science open standards and of the modular architecture of the platform, to develop new plug in modules to analyze the data.

CONCLUSION

We have implemented a genome annotation system consisting of two distributed and independent components which cooperate, one managing protein sequences and the other managing contig sequences. This tool is currently deployed in a number of laboratories throughout INRA where it is used to annotate microbial genomes.

The general philosophy of the AGMIAL platform is that human experts are central to the process of annotation, the role of computers is to assist them in this complex task. Hence, the two aspects of the platform are, on one hand the

automation of the maximum number of tasks that do not require human expertise, and on the other hand the strong emphasis put on the man-machine interface that is intended to help annotators to interact efficiently with the system.

From a computer science viewpoint, the system is built from an open community of distributed and independent components, which can cooperate together. It is thus highly modular, facilitating the integration of new tools. The components are written in Java. The system is based on well established standard computer science technologies (Web services, relational database management systems, Java, etc.) and only integrates open source software allowing us to distribute the platform freely under a GNU public license.

From a user's viewpoint, besides the characteristics mentioned above, the system has, we believe, several interesting features. It is able to handle draft sequences at various level of completion. It permits a collaborative annotation from members of a laboratory by letting them organize the annotation process as they wish and provides a history mechanisms. The latter permits the tracking of all changes occurring at the gene or protein level during the annotation process. The system enforces the use of a functional classification (that can be chosen by the team of biologists at the beginning of the project).

SYSTEM REQUIREMENTS

Being written in Java, the platform can be installed on machines running under different operating systems. So far, we have only tested the platform deployment on Linux machines or a cluster of Linux machines. On the other hand, annotators do not need to be concerned about the type of machine on which the platform is running since they only interact with the system through Web interfaces and Java applets. Applets are programs designed to be executed from within a Web browser that permit the design of dynamic Web pages, enhancing the interaction of the user with the system. Annotators can thus remain in their favorite environment (Windows, Mac OS and Linux) but still interact easily and transparently with the platform running on a remote server.

AVAILABILITY

A public version of the annotated genome of *L.sakei* can be browsed at the following URL: <http://genome.jouy.inra.fr/sakei-agmial>. A demo version ('sandbox') of the AGMIAL platform is available at the URL: <http://genome.jouy.inra.fr/demo-agmial>. The complete system, (i.e. the framework) can be downloaded at the URL: <http://genome.jouy.inra.fr/agmial> and installed locally under the GNU Public License.

Groups interested in annotating newly sequenced genomes with the AGMIAL platform but lacking the manpower to install it locally can contact J.-F. Gibrat (gibrat@jouy.inra.fr) to ask for their data to be analyzed and managed on our machines.

ACKNOWLEDGEMENTS

The authors are grateful to INRA for funding this work by an award of a Postdoctoral Fellowship to K.B. and R.B., and also

to the European Union for awarding an Individual Marie-Curie Fellowship to K.B. The authors thank M. Zagorec for critical reading of the manuscript. Funding to pay the Open Access publication charges for this article was provided by INRA.

Conflict of interest statement. None declared.

REFERENCES

- Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P. *et al.* (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, **390**, 249–256.
- Stein, L. (2001) Genome annotation: from sequence to biology. *Nature Rev. Genet.*, **2**, 493–503.
- Kitano, H. (2002) Systems biology: a brief overview. *Science*, **295**, 1662–1664.
- Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. and Yuan, Y. (1998) Predicting function: from genes to genomes and back. *J. Mol. Biol.*, **283**, 707–725.
- Jefery, C.J. (2000) Moonlighting proteins. *Trends Biochem. Sci.*, **24**, 8–11.
- Riley, M. (1997) Genes and proteins in *Escherichia coli* k-12 (GenProtEC). *Nucleic Acids Res.*, **25**, 51–52.
- Galperin, M.Y. and Koonin, E.V. (1998) Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operon disruption. *In Silico Biol.*, **1**, 55–67.
- Letunic, I., Copley, R.R., Schmidt, S., Ciccarelli, F.D., Doerks, T., Schultz, J., Ponting, C.P. and Bork, P. (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res.*, **32**, D142–D144.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Pearson, W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Meth. Mol. Biol.*, **132**, 185–219.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic Acids Res.*, **30**, 276–280.
- Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C.J., Hofmann, K. and Bairoch, A. (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res.*, **30**, 235–238.
- Marin, A., Pothier, J., Zimmermann, K. and Gibrat, J.-F. (2002) FROST: a filterbased fold recognition method. *Proteins*, **49**, 493–509.
- Notebaart, R.A., Huynen, M.A., Teusink, B., Siezen, R.J. and Snel, B. (2005) Correlation between sequence conservation and the genomic context after gene duplication. *Nucleic Acids Res.*, **33**, 6164–6171.
- Huynen, M.A., Snel, B., Mering, C. and Bork, P. (2003) Function prediction and protein networks. *Curr. Opin. Cell Biol.*, **15**, 191–198.
- Gardy, J.L., Laird, M.R., Chen, F., Rey, S., Walsh, C.J., Ester, M. and Brinkman, F.S. (2005) PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, **21**, 617–623.
- Riley, M. (1998) Systems for categorizing functions of gene products. *Curr. Opin. Struct. Biol.*, **8**, 388–392.
- The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genet.*, **25**, 25–29.
- Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M. and Mewes, H.W. (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.*, **32**, 5539–5545.
- Nicolas, P., Bize, L., Muri, F., Hoebeke, M., Rodolphe, F., Ehrlich, S.D., Prum, B. and Bessieres, P. (2002) Mining *Bacillus subtilis* chromosome heterogeneities using Hidden Markov Models. *Nucleic Acids Res.*, **30**, 1418–1426.
- Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
- d'Aubenton, C.Y., Brody, E. and Thermes, C. (1990) Prediction of rho-independent *Escherichia coli* transcription terminators: a statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.*, **216**, 835–858.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A. and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics*, **16**, 944–945.
- Hoebeke, M., Nicolas, P. and Bessieres, P. (2003) MuGeN: simultaneous exploration of multiple genomes and computer analysis results. *Bioinformatics*, **19**, 859–864.
- Stothard, P. and Wishart, D.S. (2005) Circular genome visualization and exploration using CGView. *Bioinformatics*, **21**, 537–539.
- Wootton, J.C. (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput. Chem.*, **18**, 269–285.
- Jones, D.T., Taylor, W.R. and Thornton, J.M. (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*, **33**, 3038–3049.
- Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
- von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.*, **14**, 4683–4690.
- Zdobnov, E.M. and Apweiler, R. (1986) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, **17**, 847–848.
- Haft, D.H., Selengut, J.D. and White, O. (2003) The TIGRFAMs database of protein families. *Nucleic Acids Res.*, **31**, 371–373.
- Bauer-Marchler, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwartz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z. *et al.* (2005) CDD: a conserved domain database for protein classification. *Nucleic Acids Res.*, **33**, D192–D196.
- Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. and Koonin, E.V. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.*, **29**, 22–28.
- Clamp, M., Cuff, J., Searle, S.M. and Barton, G.J. (2004) The Jalview java alignment editor. *Bioinformatics*, **20**, 426–427.
- Attwood, T.K., Bradley, P., Flower, D.R., Gaulton, A., Maudling, N., Mitchell, A.L., Moulton, G., Nordle, A., Paine, K. and Taylor, P. (2003) PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.*, **31**, 400–402.
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
- Chaillou, S., Champomier-Vergs, M.-C., Cornet, M., Crutz-Le Coq, A.-M., Dudez, A.-M., Martin, V., Beaufils, S., Darbon-Rongre, E., Bossy, R. and Loux, V. (2005) The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23k. *Nat. Biotechnol.*, **23**, 1527–1533.
- van de Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C., Oztas, S., Mangenot, S., Couloux, A. *et al.* (2006) The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc. Natl Acad. Sci., USA*, **103**, 9274–9279.
- Gaasterland, T. and Sensen, C.W. (1996) Fully automated genome analysis that reflects user needs and preferences. a detailed introduction to the MAGPIE system architecture. *Biochimie*, **78**, 302–310.
- Harris, N.L. (1997) Genotator: a workbench for sequence annotation. *Genome Res.*, **7**, 754–762.
- Bailey, L.C., Fischer, S., Schug, J., Crabtree, J., Gibson, M. and Overton, G.C. (1998) GAIA: framework annotation of genomic sequence. *Genome Res.*, **8**, 234–250.
- Medigue, C., Rechenmann, F., Danchin, A. and Viari, A. (1999) Imagene: an integrated computer environment for sequence annotation and analysis. *Bioinformatics*, **15**, 2–15.
- Andrade, M.A., Brown, N.P., Leroy, C., Hoersch, S., de Daruvar, A., Reich, C., Franchini, A., Tamames, J., Valencia, A. and Ouzounis, C. (1999) Automated genome sequence analysis and annotation. *Bioinformatics*, **15**, 391–412.
- Decker, K., Zheng, X. and Schmidt, C. (2001) A multi-agent system for automated genomic annotation. In *Proceedings of the fifth international conference on Autonomous agents*, 433–440, ACM Press.

46. Frishman,D., Albermann,K., Hani,J., Heumann,K., Metanomski,A., Zollner,A. and Mewes,H.W. (2001) Functional and structural genomics using PEDANT. *Bioinformatics*, **17**, 44–57.
47. Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,L., Clark,Y., Cox,T., Cu,J., Curwen,V. and Down,T. (2002) The ENSEMBL genome database project. *Nucleic Acids Res.*, **30**, 38–41.
48. Lewis,S.E., Searle,S.M., Harris,N., Gibson,M., Lyer,V., Richter,J., Wiel,C., Bayraktaroglu,L., Birney,E. and Crosby,M.A. (2002) Apollo: a sequence annotation editor. *Genome Biol.*, **3**, 1–14.
49. Searle,S.M., Gilbert,J., Iyer,V. and Clamp,M. (2004) The OTTER annotation system. *Genome Res.*, **14**, 963–970.
50. Sakata,K., Nagamura,Y., Numa,H., Antonio,B.A., Nagasaki,H., Itonuma,A., Watanabe,W., Shimizu,Y., Horiuchi,I., Matsumoto,T., Sasaki,T. and Higo,K. (2002) RiceGAAS: an automated annotation system and database for rice genome sequence. *Nucleic Acids Res.*, **30**, 98–102.
51. Wilkinson,M.D., Block,D. and Crosby,W.L. (2002) Genquire: genome annotation browser/editor. *Bioinformatics*, **18**, 1398–1399.
52. Bazzan,A.L.C., Duarte,R., Pitinga,A.N., Schroeder,L.F., Souto,F.D.A. and Ceroni da Silva,S. (2003) ATUGC—an agent-based environment for automatic annotation of genomes. *International Journal of Cooperative Information Systems*, **12**, 241–273.
53. Meyer,F., Goesmann,A., McHardy,A.C., Bartels,D., Bekel,T., Clausen,J., Kalinowski,J., Linke,B., Rupp,O. and Giegerich,R. (2003) GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.*, **31**, 2187–2195.
54. Glasner,J.D., Liss,P., Plunkett,G., Darling,A., Prasad,T., Rusch,M., Byrnes,A., Gilson,M., Biehler,B. and Blattner,F.R. (2002) ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.*, **31**, 147–151.
55. Almeida,L.G.P., Paixao,R., Souza,R.C., da Costa,G.C., Barrientos,F.J., dos Santos,M.T., de Almeida,D.F. and Vasconcelos,A.T.R. (2004) A system for automated bacterial (genome) integrated annotation-SABIA. *Bioinformatics*, **20**, 2832–2833.
56. Vallenet,D., Labarre,L., Rouy,Z., Barbe,V., Bocs,S., Cruveiller,S., Lajus,A., Pascal,G., Scarpelli,C. and Medigue,C. (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.*, **34**, 53–65.
57. Gattiker,A., Gasteiger,E. and Bairoch,A. (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics*, **1**, 107–108.
58. Sigrist,C.J.A., Cerutti,L., Hulo,N., Gattiker,A., Falquet,L., Pagni,M., Bairoch,A. and Bucher,P. (2002) PROSITE: a documented data base using patterns and profiles as motif descriptors. *Brief Bioinform*, **3**, 265–274.
59. Scordis,P., Flower,D.R. and Attwood,T.K. (1999) FingerPRINTSscan: intelligent searching of the PRINTS motif database. *Bioinformatics*, **15**, 799–806.
60. Wu,C.H., Nikolskaya,A., Huang,H., Yeh,L.S., Natale,D.A., Vinayaka,C.R., Hu,Z.Z., Mazumder,R., Kumar,S. and Kourtesis,P. (2004) PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Res.*, **32**, D112–D114.
61. Mi,H., Lazareva-Ulitsky,B., Loo,R., Kejariwal,A., Vandergriff,J., Rabkin,S., Guo,N., Muruganujan,A., Doremiex,O. and Campbell,M.J. (2005) The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.*, **33**, D284–D288.
62. Gough,J., Krupar,K., Hughey,R. and Chothia,C. (2001) Assignment of homology to genome sequences using a library of Hidden Markov Models that represent all proteins of known structure. *J. Mol. Biol.*, **313**, 903–919.
63. Lee,D., Grant,A., Marsden,R.L. and Orengo,C. (2005) Identification and distribution of protein families in 120 completed genomes using Gene3D. *Proteins*, **59**, 603–615.
64. Bru,C., Courcelle,E., Carrere,S., Beausse,Y., Dalmar,S. and Kahn,D. (2005) The ProDom database of protein domain families: more emphasis on 3d. *Nucleic Acids Res.*, **33**, D212–D215.
65. Dyrlov Bendtsen,J., Nielsen,H., von Heijne,G. and Brunak,S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, **340**, 783–795.
66. Moller,S., Croning,M.D.R. and Apweiler,R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **17**, 646–653.
67. Bairoch,A., Apweiler,R., Wu,C.H., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R. and Magrane,M. (2005) The universal protein resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
68. Deshpande,N., Address,K.J., Bluhm,W.F., Merino-Ott,J.C., Townsend-Merino,W., Zhang,Q., Knezevich,C., Xie,L., Chen,L. and Feng,L. (2005) The RCSB protein data bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res.*, **33**, D233–D237.
69. Andreeva,A., Howorth,D., Brenner,S.E., Hubbard,T.J., Chothia,C. and Murzin,A.G. (2004) SCOP database in 2004: recent improvements integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226–D229.
70. Hulo,N., Sigrist,C.J., Le Saux,V., Langendijk-Genevaux,P.S., Bordoli,L., Gattiker,A., De Castro,E., Bucher,P. and Bairoch,A. (2004) Recent improvements to the PROSITE database. *Nucleic Acids Res.*, **32**, D134–D137.
71. Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths,J., Khanna,A., Marshall,M., Moxon,S. and Sonnhammer,E.L. (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
72. Haft,D.H., Selengut,J.D. and White,O. (2003) The TIGRFAMs database of protein families. *Nucleic Acids Res.*, **31**, 371–377.

Le travail décrit dans cette thèse a commencé à un moment où peu de données génétiques sur *Lactobacillus delbrueckii* ssp. *bulgaricus* (*L. bulgaricus*) étaient disponibles. Le séquençage du génome de la souche ATCC 11842, en collaboration avec le Génoscope, était en cours. La première partie de cette thèse décrit le travail effectué sur l'annotation et l'analyse du génome. Cette dernière suggère que *L. bulgaricus* est en voie de spécialisation à une niche écologique restreinte, le milieu lait. Elle révèle également que *L. bulgaricus* possède peu de systèmes de régulation et de résistance aux stress environnementaux. Ces résultats nous ont conduit à faire l'hypothèse que *L. bulgaricus* serait un bon modèle pour identifier les mécanismes essentiels pour l'adaptation à l'acidité (ATR).

La deuxième partie de la thèse décrit les études menées sur l'ATR de *L. bulgaricus* par différentes approches complémentaires. Les résultats transcriptomiques mettent en évidence que l'expression de 20% des régulateurs ainsi que celle de 50% des transporteurs d'ions de *L. bulgaricus* est modifiée par l'ATR. En revanche, aucun des 5 facteurs σ ni des 5 systèmes à deux composants du génome ne présente une expression variable. L'analyse protéomique par électrophorèse bidimensionnelle révèle 19 protéines dont l'abondance est affectée par l'ATR, parmi lesquelles les protéines chaperons et des protéines impliquées dans la biosynthèse des acides gras. L'ensemble des résultats suggère que l'ATR de *L. bulgaricus* fait intervenir de multiples régulateurs, des phénomènes d'atténuation de la transcription, et des régulations post-transcriptionnelles et post-traductionnelles. Ces résultats préliminaires nécessitent cependant des analyses complémentaires pour déterminer quelle partie des réponses visualisées est réellement à la base de la résistance accrue à l'acidité acquise après adaptation.

The work described here began at a moment when few genetic data about *Lactobacillus delbrueckii* ssp. *bulgaricus* (*L. bulgaricus*) were available. Genome of the strain ATCC 11842, in collaboration with the Genoscope, was being sequenced. The first part of this thesis describes the genome annotation and analyse. This analyse suggest that *L. bulgaricus* is going on a specialisation to a restricted ecological area, the milk medium. It also reveals that *L. bulgaricus* has few regulation systems and resistance mechanisms to environmental stresses. This results led us to make the hypothesis that *L. bulgaricus* is a good model to identify important mechanisms for the acid tolerance response (ATR).

The second part of the thesis describes the studies carried on the ATR of *L. bulgaricus* by complementary approaches. Transcriptomic results show that expression of 20% of the regulators and 50% of the ion transporters of the genome was modified by ATR. Nevertheless, none of the 5 σ factor nor of the 5 two-component system present a variable expression. The proteomic analyse show that 19 proteins have their amount affected by ATR, among which chaperones and proteins involved in fatty acids biosynthesis. All together, our results suggest that ATR of *L. bulgaricus* needs several regulators, transcriptional attenuation mechanisms and post-transcriptional and post-traductional regulations. These preliminary results needs complementary analyses to determine which part of the visualised responses is really necessary for the increased resistance to acid stress after adaptation.